Redshift distributions of galaxies in the Dark Energy Survey Science Verification shear catalogue and implications for weak lensing

C. Bonnett,[1] M. A. Troxel,[2] W. Hartley,[3] A. Amara,[3] B. Leistedt,[4] M. R. Becker,[5,6] G. M. Bernstein,[7] S. L. Bridle,[2] C. Bruderer,[3] M. T. Busha,[5] M. Carrasco Kind,[9,10] M. J. Childress,[48] F. J. Castander,[11] C. Chang,[3] M. Crocce,[11] T. M. Davis,[47] T. F. Eifler,[7,12] J. Frieman,[13,14] C. Gangkofner,[51,52] E. Gaztanaga,[11] K. Glazebrook,[49] D. Gruen,[33,35] T. Kacprzak,[3] A. King,[8] J. Kwan,[15] O. Lahav,[4] G. Lewis,[46] C. Lidman,[16] H. Lin,[13] N. MacCrann,[2] R. Miquel,[1,17] C. R. O'Neill,[47] A. Palmese,[4] H. V. Peiris,[4] A. Refregier,[3] E. Rozo,[18] E. S. Rykoff,[6,19] I. Sadeh,[4] C. Sánchez,[1] E. Sheldon,[50] S. Uddin,[20] R. H. Wechsler,[5,6,19] J. Zuntz,[2] T. Abbott,[21] F. B. Abdalla,[4] S. Allam,[13] R. Armstrong,[22] M. Banerji,[23,24] A. H. Bauer,[11] A. Benoit-Lévy,[4] E. Bertin,[25,26] D. Brooks,[4] E. Buckley-Geer,[13] D. L. Burke,[6,19] D. Capozzi,[27] A. Carnero Rosell,[28,29] J. Carretero,[1,11] C. E. Cunha,[6] C. B. D'Andrea,[27] L. N. da Costa,[28,29] D. L. DePoy,[30] S. Desai,[31,32] H. T. Diehl,[13] J. P. Dietrich,[32,33] P. Doel,[4] A. Fausti Neto,[28] E. Fernandez,[1] B. Flaugher,[13] P. Fosalba,[11] D. W. Gerdes,[34] R. A. Gruendl,[9,10] K. Honscheid,[36,37] B. Jain,[7] D. J. James,[21] M. Jarvis,[7] A. G. Kim,[38] K. Kuehn,[16] N. Kuropatkin,[13] T. S. Li,[30] M. Lima,[45,28] M. A. G. Maia,[28,29] M. March,[7] J. L. Marshall,[30] P. Martini,[36,39] P. Melchior,[36,37] C. J. Miller,[34,40] E. Neilsen,[13] R. C. Nichol,[27] B. Nord,[13] R. Ogando,[28,29] A. A. Plazas,[12] K. Reil,[19] A. K. Romer,[41] A. Roodman,[6,19] M. Sako,[7] E. Sanchez,[42] B. Santiago,[28,43] R. C. Smith,[21] M. Soares-Santos,[13] F. Sobreira,[13,28] E. Suchyta,[36,37] M. E. C. Swanson,[10] G. Tarle,[34] J. Thaler,[44] D. Thomas,[27] V. Vikram,[15] and A. R. Walker[21]

(Dark Energy Survey Collaboration)

[1]*Institut de Física d'Altes Energies, Universitat Autònoma de Barcelona, E-08193 Bellaterra, Barcelona, Spain*
[2]*Jodrell Bank Center for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom*
[3]*Department of Physics, ETH Zurich, Wolfgang-Pauli-Strasse 16, CH-8093 Zurich, Switzerland*
[4]*Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, United Kingdom*
[5]*Department of Physics, Stanford University, 382 Via Pueblo Mall, Stanford, California 94305, USA*
[6]*Kavli Institute for Particle Astrophysics & Cosmology, P. O. Box 2450, Stanford University, Stanford, California 94305, USA*
[7]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA*
[8]*School of Mathematics and Physics, University of Queensland, Brisbane, QLD 4072, Australia*
[9]*Department of Astronomy, University of Illinois, 1002 W. Green Street, Urbana, Illinois 61801, USA*
[10]*National Center for Supercomputing Applications, 1205 West Clark St., Urbana, Illinois 61801, USA*
[11]*Institut de Ciències de l'Espai, IEEC-CSIC, Campus UAB, Carrer de Can Magrans, s/n, 08193 Bellaterra, Barcelona, Spain*
[12]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, California 91109, USA*
[13]*Fermi National Accelerator Laboratory, P. O. Box 500, Batavia, Illinois 60510, USA*
[14]*Kavli Institute for Cosmological Physics, University of Chicago, Chicago, Illinois 60637, USA*
[15]*Argonne National Laboratory, 9700 South Cass Avenue, Lemont, Illinois 60439, USA*
[16]*Australian Astronomical Observatory, North Ryde, NSW 2113, Australia*
[17]*Institució Catalana de Recerca i Estudis Avançats, E-08010 Barcelona, Spain*
[18]*Department of Physics, University of Arizona, Tucson, Arizona 85721, USA*
[19]*SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA*
[20]*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Victoria 3122, Australia*
[21]*Cerro Tololo Inter-American Observatory, National Optical Astronomy Observatory, Casilla 603 La Serena, Chile*
[22]*Department of Astrophysical Sciences, Princeton University, Peyton Hall, Princeton, New Jersey 08544, USA*
[23]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*
[24]*Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom*
[25]*CNRS, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*
[26]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7095, Institut d'Astrophysique de Paris, F-75014 Paris, France*

[27]*Institute of Cosmology & Gravitation, University of Portsmouth, Portsmouth PO1 3FX, United Kingdom*
[28]*Laboratório Interinstitucional de e-Astronomia - LIneA, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[29]*Observatório Nacional, Rua Gal. José Cristino 77, Rio de Janeiro, RJ 20921-400, Brazil*
[30]*George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, and Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843, USA*
[31]*Department of Physics, Ludwig-Maximilians-Universität, Scheinerstr. 1, 81679 München, Germany*
[32]*Excellence Cluster Universe, Boltzmannstr. 2, 85748 Garching, Germany*
[33]*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, 81679 München, Germany*
[34]*Department of Physics, University of Michigan, Ann Arbor, Michigan 48109, USA*
[35]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748 Garching, Germany*
[36]*Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[37]*Department of Physics, The Ohio State University, Columbus, Ohio 43210, USA*
[38]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*
[39]*Department of Astronomy, The Ohio State University, Columbus, Ohio 43210, USA*
[40]*Department of Astronomy, University of Michigan, Ann Arbor, Michigan 48109, USA*
[41]*Department of Physics and Astronomy, Pevensey Building, University of Sussex, Brighton BN1 9QH, United Kingdom*
[42]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[43]*Instituto de Física, UFRGS, Caixa Postal 15051, Porto Alegre, RS 91501-970, Brazil*
[44]*Department of Physics, University of Illinois, 1110 W. Green St., Urbana, Illinois 61801, USA*
[45]*Departamento de Física Matemática, Instituto de Física, Universidade de São Paulo, CP 66318, CEP 05314-970 São Paulo, SP, Brazil*
[46]*SEPnet, South East Physics Network, (www.sepnet.ac.uk)*
[47]*School of Mathematics and Physics, University of Queensland, QLD 4072, Australia*
[48]*Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611, Australia*
[49]*Centre for Astrophysics & Supercomputing, Swinburne University of Technology, Victoria 3122, Australia*
[50]*Brookhaven National Laboratory, Bldg 510, Upton, New York 11973, USA*
[51]*Excellence Cluster Universe, Boltzmannstr. 2, 85748 Garching, Germany*
[52]*Faculty of Physics, Ludwig-Maximilians University, Scheinerstr. 1, 81679 Munich, Germany*
(Received 31 July 2015; published 30 August 2016)

We present photometric redshift estimates for galaxies used in the weak lensing analysis of the Dark Energy Survey Science Verification (DES SV) data. Four model- or machine learning-based photometric redshift methods—ANNZ2, BPZ calibrated against BCC-Ufig simulations, SKYNET, and TPZ—are analyzed. For training, calibration, and testing of these methods, we construct a catalogue of spectroscopically confirmed galaxies matched against DES SV data. The performance of the methods is evaluated against the matched spectroscopic catalogue, focusing on metrics relevant for weak lensing analyses, with additional validation against COSMOS photo-z's. From the galaxies in the DES SV shear catalogue, which have mean redshift 0.72 ± 0.01 over the range $0.3 < z < 1.3$, we construct three tomographic bins with means of $z = \{0.45, 0.67, 1.00\}$. These bins each have systematic uncertainties $\delta z \lesssim 0.05$ in the mean of the fiducial SKYNET photo-z $n(z)$. We propagate the errors in the redshift distributions through to their impact on cosmological parameters estimated with cosmic shear, and find that they cause shifts in the value of σ_8 of approximately 3%. This shift is within the one sigma statistical errors on σ_8 for the DES SV shear catalogue. We further study the potential impact of systematic differences on the critical surface density, $\Sigma_{\rm crit}$, finding levels of bias safely less than the statistical power of DES SV data. We recommend a final Gaussian prior for the photo-z bias in the mean of $n(z)$ of width 0.05 for each of the three tomographic bins, and show that this is a sufficient bias model for the corresponding cosmology analysis.

DOI: 10.1103/PhysRevD.94.042005

I. INTRODUCTION

One of the key goals of the Dark Energy Survey (DES) is to extract cosmological information from measurements of weak gravitational lensing. Gravitational lensing (for discussion see Refs. [1–3], and references therein) involves the deflection of light from distant galaxies by intervening matter along the line of sight. Lensing encodes information in the shapes of background objects (i.e., galaxies) on both the statistical properties of intervening matter perturbations and cosmological distances to the sources. The primary challenge in studying gravitational lensing in the weak regime has been the difficulty in measuring the shapes of galaxies in an unbiased way. For a detailed discussion of galaxy shape measurements in DES Science Verification (DES SV), see Ref. [4]. However, a weak lensing analysis requires not only the careful measurement of the shapes of galaxies, but also an accurate and unbiased estimate of redshifts to a large ensemble of galaxies.

Knowing the redshifts of the galaxies in a sample (or equivalently, their distances for a given cosmological model), allows us to differentiate near and distant galaxies and thereby reconstruct the redshift dependence of the lensing signal. Hence separating galaxies into redshift bins strongly improves the constraining power of cosmic shear on cosmological model parameters [5]. Extensive studies have been reported in the literature that look for optimal configurations of redshift binning and requirements for future ambitious surveys, covering several thousand square degrees, [6–17]. In addition to gains in statistical precision, separating galaxies into tomographic bins can also mitigate astrophysical systematics. For example, moving to a tomographic analysis allows us to better isolate the intrinsic correlations of galaxy shapes in the absence of lensing (see Refs. [18,19] and references therein), whereas a nontomographic analysis may otherwise be limited by uncertainties in the impact of this intrinsic galaxy alignment (for more, see Ref. [20]).

Given the large number of galaxies that make up a lensing sample in a wide field imaging survey, redshifts must be estimated using photometry measured in a series of (typically) broad bands. This method of estimating photometric redshifts is known as photo-z (see Ref. [21], and discussion and references therein). Achieving the high level of precision necessary to ensure that the systematic contributions to cosmological parameter uncertainties due to photo-z bias are of the order of the statistical uncertainty is challenging, as is the necessary validation of the derived redshifts [22–27]. Previous weak lensing surveys have tackled this problem in a variety of innovative ways. For example, see Refs. [24] and [23] for the discussion of this problem in the CFHTLenS survey [28,29] and Ref. [27] in the Deep Lens Survey [30]. Substantial and dedicated efforts are required to improve current performance and achieve the target precision in on-going and future surveys. The challenging target set for the full Dark Energy Survey is that the biases in redshift estimates of the means of tomographic bins should be below $\delta z = 0.003$, which is based on the desire to keep redshift systematic errors subdominant to the statistical errors of the lensing surveys [6,7].

In this work we explore accurate and precisely characterized photo-z estimates of $n(z)$, the result of stacking the individual probability distribution functions $p(z)$, with the SV data of DES. At 139 square degrees, the required precision for DES SV weak lensing analyses are significantly weaker than those for the full DES survey data. As such we target precision at the few percent level for the mean redshifts of a given population of galaxies. This will allow us to have photo-z uncertainties comparable to or lower than the statistical errors on the cosmological parameters we are best able to constrain (e.g., σ_8). We can study the impact of redshift precision directly by propagating the expected photo-z bias to the constraints on σ_8, but also by comparing the differences in final predictions for σ_8 over the full DES SV shear catalogue from each of four different independent photometric redshift methods.

The paper is organized as follows. In Sec. II we introduce the data products that are used in our studies. In Secs. III and IV we investigate the global properties of the lensing sample including magnitude, color and redshift distributions. We also discuss the limitations of existing spectroscopic samples. In Sec. V we extend our analysis to tomographic cases and the impact on cosmological parameters is explored in Sec. VI. Our conclusions are summarized in Sec. VII.

II. DATA SETS

Prior to the start of the main Dark Energy Survey, the Dark Energy Camera (DECam) [31–34], with a hexagonal footprint of 570 megapixels, was tested during a preliminary SV survey from November 2012 to February 2013. These observations produced a useable DES SV galaxy catalogue with which measurement and analysis pipelines have been tested to produce early science results. The DES SV survey mimics full 5-year DES survey parameters over a small patch of the sky, but with significant depth variations due to weather and other challenges during early operations of DECam (see e.g., Ref. [35]). The contiguous area used for the DES SV shear catalogue is contained within the South Pole Telescope east (SPT-E) observing region [36], and covers approximately 139 square degrees in five optical filters, g, r, i, z, and Y. We note that the Y band was not used in this work.

In this section we present the DES SV data products relevant for photometric redshift estimation. We also build a catalogue of precise and reliable spectroscopic redshifts by collating a number of proprietary and public spectroscopic data sets that also have DES photometric observations available. This is essential to test the methods for

photo-z estimates used in this work. Finally, we describe a set of simulations of the DES SV survey that we use as a secondary method of calibrating and validating the photo-z estimates.

A. DES SV photometry and Gold catalogue

DES data from the SV season were reduced by the SVA1 version of the DES Data Management system [37], using SCAMP [38], SWARP [39] and bespoke software packages, as described in Refs. [40,41] and [37]. To summarize, the single-epoch images were calibrated, background-subtracted, coadded, and processed in "tiles" (0.75×0.75 deg^2 squares) defined to cover the entire DES footprint. A catalogue of objects was extracted from the coadded images using SOURCE EXTRACTOR (SEXTRACTOR, [42,43]). In what follows we use *AB* magnitudes and MAG_AUTO measurements performed in coadd images, which are reliable for SV galaxies (e.g., robust to sharp point spread function variations across coadd images) and used in most SV analyses (e.g., Ref. [44]). However, note that shape measurements are performed in single-epoch images with a dedicated pipeline using multi-epoch fitting techniques, as described in Ref. [4]. The analysis presented in this work will be concerned with the objects that meet the quality cuts of that pipeline.

The main catalogue of reliable objects in DES SV is the Gold catalogue described in Ref. [45]. It starts with all objects detected in SV images and successively applies quality cuts to reject objects and regions that are deemed problematic (e.g., regions with poor observations or photometry). To be included in the Gold catalogue, an object must

(i) be observed at least once in all four *griz* bands,
(ii) be at a declination above $-61°$ to avoid regions of bad photometric calibration (e.g., Large Magellanic Cloud),
(iii) not be in regions with galaxy surface density $> 3\sigma$ below the mean,
(iv) not be in regions surrounding bright stars, and
(v) not be in regions with a concentration of large centroid shifts or dropouts between bandpasses.

Further information on star-galaxy separation and quality cuts at the shape measurement level are described in detail in Ref. [4].

B. DES SV shear catalogue

Two semi-independent shear pipelines—IM3SHAPE and NGMIX—have been produced for a subset of objects in the DES SV Gold catalogue in the SPT-E region of the sky. These are described further in Ref. [4], but relevant details are summarized below. The two shear pipelines produce separate shear measurements for each galaxy, and thus select a different subset of the galaxies in the Gold catalogue as having well-measured shears. This leads to a different population of galaxies used by either pipeline in constructing the $n(z)$ for each tomographic bin in a weak lensing analysis, though the IM3SHAPE selection is nearly a subset of the NGMIX selection. The final shear catalogue is the intersection of the Gold galaxy selection, these shear-related cuts, and a final "good" galaxy selection for lensing that removes objects with SEXTRACTOR flags $= 1, 2$, much lower surface-brightness objects, very small objects, or those with colors outside reasonable bounds ($-1 < g - r < 4$ and $-1 < i - z < 4$). These selection effects also produce slightly different photometric properties in the galaxy sample used. This is demonstrated in Fig. 1, where the i-mag histogram is compared for all "Gold" objects, all galaxies, "good" galaxies, as defined above, and finally the two shear selections.

(i) *IM3SHAPE*: The IM3SHAPE shear measurement pipeline is built on the IM3SHAPE code discussed in Ref. [46] and modified as described in Ref. [4]. The IM3SHAPE code is a forward-modeling maximum-likelihood method that fits two galaxy models to an image in the r band: an exponential disc and a de Vaucouleurs bulge. The best-fitting model is then used to estimate the ellipticity. Inverse variance weights are calculated for each galaxy empirically in bins of size and signal to noise. The final IM3SHAPE shear catalogue has a number density of $\simeq$4.2 galaxies per square arcminute.
(ii) *NGMIX*: The NGMIX shear measurement pipeline represents simple galaxy models as the sum of Gaussians [47]. The same model shape is fit simultaneously across the *riz* bands, with parameters sampled via Markov chain Monte Carlo techniques. Ellipticities are then estimated using the LENSFIT algorithm [48] with priors on the intrinsic ellipticity distribution from GREAT3 [49]. Inverse variance weights are calculated for each galaxy from the covariance of the shape estimate and an intrinsic shape noise estimate. The final NGMIX shear catalogue has a number density of $\simeq$6.9 galaxies per square arcminute.

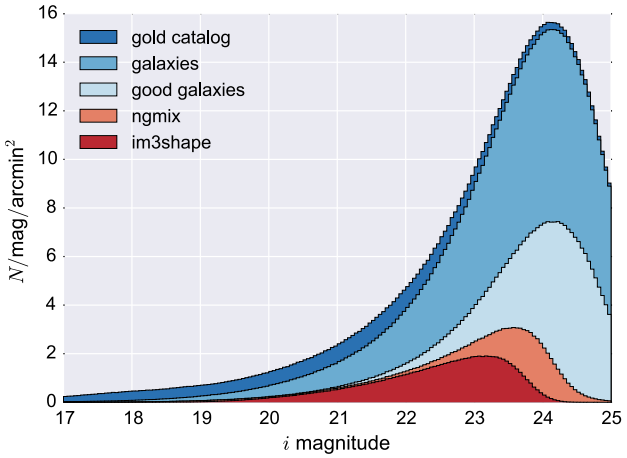

FIG. 1. i-band magnitude histograms for various levels of cuts from the full Gold catalogue down to the final shear catalogue.

Throughout this work we use the NGMIX catalogue as the default weak lensing sample unless explicitly stated otherwise.

C. Spectroscopic catalogues

To train and assess the performance of the photometric redshifts we assemble a matched catalogue of galaxies that are observed with both DECam and a spectrograph. In this section we describe the photometric and spectroscopic properties of this matched catalogue. Objects are matched on the sky within a matching radius of 1.5 arcseconds. The spectra used come from six distinct areas on the sky and contain a total of 46 139 galaxies. The distributions of these fields on the sky relative to the main DES SV SPT-E field are shown in Fig. 2. In Table I the general properties of the spectroscopic surveys used in this matched catalogue are listed, but for a more detailed description of the properties (e.g., the quality flags used), we refer the reader to Appendix A. We note that the combination of VVDS-F02 Deep, VVDS CDFS Deep and VVDS Ultra Deep is referred to as VVDS-Deep in rest of this work.

The final matched spectroscopic catalogue has been cleaned of objects that we do not expect to be present in the shear catalogue. This includes removing all stars, strong lenses, and active galactic nuclei (AGN). The matching is limited to the $(0 < z < 1.8)$ redshift range. This means that for all the machine learning (ML) methods used in this work the density of $n(z)$ above $z = 1.8$ will be zero, though model-fitting codes do not have this drawback. We test that artificially cutting the $n(z)$ at 1.8 for a model-fitting code biases the constraints on σ_8 at the 1% level, which is sufficiently small relative to the statistical error (see Sec. VI for more details).

We divide the resulting matched spectroscopic catalogue into three samples: a training, a validation, and an independent sample, which are compared in Fig. 3. The independent sample contains all the matched galaxies from the VVDS-F14 field; a total of 3603 galaxies. This field is spatially removed from the other spectroscopic fields, as shown in Fig. 2, and therefore the line of sight structure within this field is uncorrelated with that of training and validation sets. The use of this field will allow us to assess issues pertaining to sample variance and radial learning in

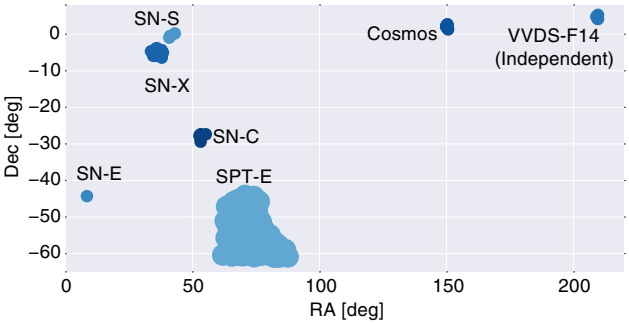

FIG. 2. Location of the six spectral fields and the main DES SV (SPT-East) field on the sky. The SN fields are the DES supernova fields while the other two have been observed with DECam outside of the DES survey.

TABLE I. The number of galaxies that are included in the matched spectroscopic catalogue are listed for each spectroscopic survey with the corresponding mean redshift and mean i band magnitude. Further details can be found in Appendix A.

Spectroscopic survey	Count	Mean i	Mean z
VIPERS	7286	21.52	0.69
GAMA	7276	18.61	0.22
Zcosmos	5442	20.93	0.51
VVDS F02 Deep	4381	22.40	0.68
SDSS	4140	18.82	0.3 9
ACES	3677	21.73	0.58
VVDS F14	3603	20.61	0.49
OzDES	3573	19.85	0.47
ELG cosmos	1278	22.22	1.08
SNLS	857	21.09	0.55
UDS VIMOS	774	22.54	0.85
2dFGRS	725	17.52	0.13
ATLAS	722	18.96	0.35
VVDS spF10 WIDE	661	21.16	0.53
VVDS CDFS DEEP	544	22.05	0.62
UDS FORS2	311	23.80	1.25
PanSTARRS MMT	297	19.94	0.35
VVDS Ultra DEEP	264	23.71	0.88
PanSTARRS AAOmega	239	19.69	0.32
SNLS AAOmega	81	21.16	0.56

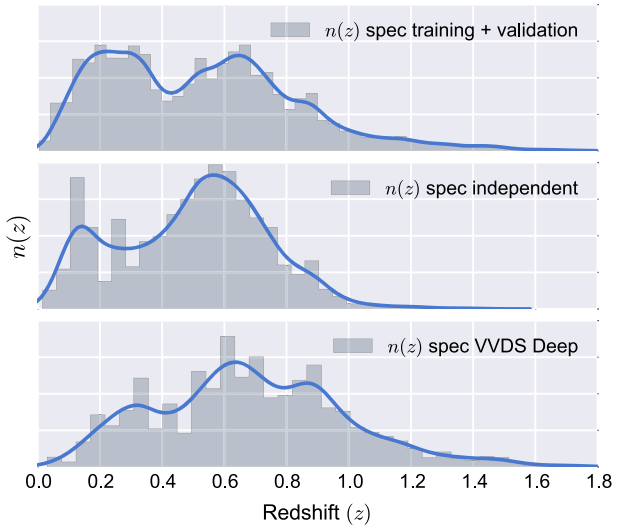

FIG. 3. The normalized redshift distributions of the spectroscopic samples used in producing and testing the photometric redshift estimates. The solid line is the kernel density estimate [50] estimate of the underlying density. Top panel: The combined training and validation samples. Middle panel: The independent sample (VVDS-F14). Bottom panel: The VVDS-Deep sample.

the machine learning methods (e.g., Appendix D). If the redshift solution is overtrained or subject to systematic incompleteness, any performance metrics on a validation set with a near identical redshift distribution to the training sample would be too optimistic. In Appendix D, we demonstrate an example of extreme selection effects in a training set based on the PRIMUS survey, while in Sec. III C we study the completeness of the training set used in this work. The remaining 42 536 galaxies in the matched spectroscopic catalogue are split into the training and validation samples containing, respectively, 70% and 30% of the galaxies. This retains a total of 28 219 galaxies in the training sample and 14 317 galaxies in the validation sample.

D. COSMOS data

In addition to spectroscopic data from the literature, we also make use of the point-estimated photometric redshifts from Ref. [51] in the COSMOS field. These photo-z estimates were computed from 30-band photometry with the Le Phare template-fitting photometric redshift code [52]. The COSMOS field was observed with DECam during the SV observing period and coadd images with a similar total exposure time as the SV survey have been produced. We match the catalogue extracted from these images to the COSMOS photo-z sample, and trim to a subsample representative of the shear catalogue. This trimming was performed by applying cuts in the i-band FWHM - magnitude plane as follows:

$$\begin{aligned} \mathrm{FWHM\,(arcsec)} &> 0.105 \times i(\mathrm{mag}) - 1, \\ \mathrm{FWHM\,(arcsec)} &> 0.751 \times i(\mathrm{mag}) - 15.63, \\ i &> 18(\mathrm{mag}), \end{aligned}$$

together with a surface brightness cut at $\mu_{\mathrm{eff}} < 28\ \mathrm{mag\,arcsec^{-2}}$. These cuts approximate the final shape catalogue selection function and allow us a further independent estimate of the redshift distribution of the weak lensing sample.

E. Simulated SV data: The BCC-UFig

In the following sections we will calibrate a model-based photo-z method using a set of galaxy catalogues extracted from simulated SV data: the BCC-UFig [53]. The latter is based on simulated DES coadd images created using the Ultra-Fast Image Simulator (UFig, [54]). The input galaxy catalogues for these images were taken from the Blind Cosmology Challenge (BCC, [55]). The galaxy catalogues were then obtained by running source extraction and processing codes to mimic the pipeline run on the real DES SV data, as described in Refs. [53] and [35]. The BCC-UFig was shown to reliably mimic the SV data in terms of color, redshift, and spatial distributions of the objects, and also reproduce systematics observed in the reduced galaxy catalogues such as spatially varying depth and correlations with observing conditions [35,53]. In this paper we push the comparison further and consider catalogues similar to the weak lensing catalogue described above by making the same catalogue-level cuts as are used for the COSMOS data.

III. PROPERTIES OF THE MATCHED SPECTROSCOPIC CATALOGUE AND TEMPLATES

Ideally, we would be able to compile a sample of spectroscopically identified objects that are fully representative of our target weak lensing galaxy population. If these spectroscopic objects were sufficiently numerous and well sampled over the sky, then the redshift distribution of these objects could be used in conjunction with weak lensing measurements to infer constraints on cosmological parameters. However, even in large samples such as the one compiled for this work, biases remain due to spectroscopic incompleteness and difficulties in representing all galaxies in the face of spatially varying data quality.

In this section we investigate to what extent our existing spectroscopic sample should reflect the underlying redshift distribution of our photometric sample and assess the effectiveness of weighting spectroscopic objects in correcting for differences between the photometric and spectroscopic galaxy populations. We pay special attention to possible biases in the inferred probability distribution of the weak lensing sources due to these limitations. Note that while modeling methods do not require representative training samples, biases may still arise if the model templates are not a sufficiently accurate description of the data. This is analogue to model bias in cosmic shear measurements [56,57]. As in cosmic shear, we can aim to tackle these issues through simulations of the data. Thus Secs. III A–III C address challenges related to machine learning methods, while Sec. III D discusses challenges to using template-fitting methods.

A. Noise properties of the matched catalogue

A large fraction of the DES-SV galaxies that have spectra lie in the DES supernovae fields or other fields with a significantly longer cumulative exposure time than the SPT-E field, which contains the galaxies used for the weak lensing science. We show in Fig. 4 the estimated 10σ MAG_AUTO detection limits of the matched spectroscopic catalogue compared to that of the weak lensing sample. The 10σ detection limits differ significantly between the samples, with the galaxies in the matched spectroscopic catalogue having significantly deeper detection limits on average. Details on the image depth calculation algorithm can be found in Ref. [58]. This poses a problem for ML methods as they do not explicitly take the noise measurement into account. The ML methods in this work implicitly

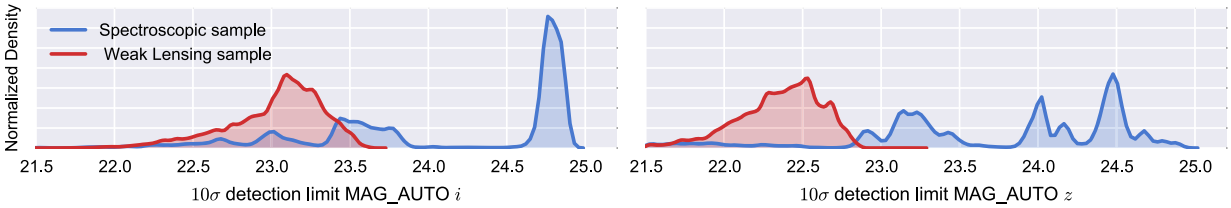

FIG. 4. The 10σ MAG_AUTO detection limits of the matched spectroscopic sample (blue) compared to that of the weak lensing sample (red). The matched spectroscopic catalogue has a significantly larger detection limit due to the fact that many DES galaxies with spectra lie in the frequently observed DES supernova fields.

assume that the noise properties from the matched spectroscopic catalogue are identical to those of the weak lensing sample.

One way to obtain a similar depth distribution in the spectroscopic set is to create coadded images of the deeper fields using a subset of exposures with numbers similar to those typical in the SPT-E field, as was used in Ref. [26]. A second option is to algorithmically degrade the photometry of the matched spectroscopic catalogue for the bands of the galaxies with higher S/N. This is done in the following manner:

(1) For every galaxy in the matched spectroscopic catalogue, we find its nearest neighbor in four-dimensional color-magnitude space, $(g-r, r-i, i-z, i)$, from the weak lensing sample.
(2) If one or more bands of the matched galaxy have a fainter 10σ detection limit than the weak lensing sample detection limit in those bands, then a new magnitude is drawn.
(3) This new magnitude is determined according to a normal distribution using the measured magnitude of the spectroscopic galaxy as the mean and the error on the magnitude of the selected neighbor in the weak lensing sample for the variance.

The limits in image depth (10σ detection) for which we decide to redraw a new magnitude value are MAG_AUTO $g = 24.5$, MAG_AUTO $r = 24.3$, MAG_AUTO $i = 23.5$, and MAG_AUTO $z = 22.8$. So, for a galaxy in the matched spectroscopic catalogue that has a 10σ detection of 24.7 in the i band and a 10σ detection of 22.5 in the z band we draw a new i band magnitude and keep the original z band magnitude.

This leads to a matched spectroscopic catalogue that has approximately the same noise properties as the weak lensing sample. The method has some advantages over restacking, one of which is that we can degrade to any other noise level as long as the original exposures are of sufficient depth. This is not necessarily possible with restacking due to the fact that observing conditions sampled during pointings in SPT-E cannot be recreated with those observed in the deeper fields. To protect against potential biases introduced by this procedure, the training and validation in this work have been algorithmically degraded while the independent field containing all the VVDS-F14 galaxies is created by restacking and is identical to the reduction of the field used in Ref. [26]. We validated that using restacked coadds instead of resampling the magnitudes has no significant effect on our results.

B. Weighting of the spectroscopic set

In the work presented here we characterize the impact of errors in redshift estimation on weak lensing studies. Our focus is thus on the galaxy samples selected based on our ability to measure accurately their shapes in DES SV. Figure 5 shows the i-band magnitude distribution of the matched spectroscopic catalogue in blue and the distribution of the weak lensing sample from DES SV in red. The difference in magnitude of the samples is very clear, with the matched spectroscopic sample biased to brighter magnitudes. We account for differences in magnitude and color by weighting galaxies in the spectroscopic sample in such a way that the weighted distribution of training galaxies matches the weak lensing source distribution. This can then be used in performance metrics to give a better indication of the likely errors coming from averaging over the weak lensing population. The weights we use are calculated as in Ref. [26] by estimating the density of objects in the matched spectroscopic sample in color-magnitude space noted below, with all objects detected in all bands, and

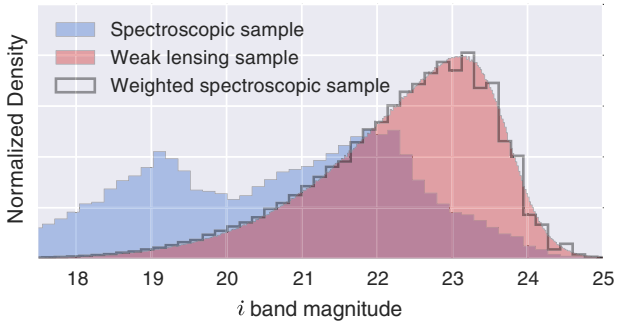

FIG. 5. The i-band magnitude distribution of the matched spectroscopic catalogue is shown in blue and the weak lensing sample is shown in red. The matched spectroscopic catalogue after weighting is shown as the grey histogram outline overlaying the weak lensing sample.

$$-1 < g - r < 4,$$
$$-1 < r - i < 4,$$
$$-1 < i - z < 4,$$
$$16 < i,$$
$$16 < r.$$

We then compare this density with the density of the weak lensing sample at the same location in color-magnitude space, using the NGMIX catalogue. The ratio of the densities of the weak lensing sample to the matched spectroscopic catalogue at the location of a spectroscopic galaxy in color-magnitude space is calculated by counting the number of galaxies in the weak lensing sample in a hypersphere with radius to the fifth nearest neighbor in Euclidian space in the matched spectroscopic catalogue. The normalized ratios of these densities are then used as weights for the spectroscopic galaxies (see Ref. [59] for more details on the implementation).

Figure 5 shows the weighted i-band distribution for the spectroscopic sample, which better matches the NGMIX catalogue. In Fig. 6, we show $g-r$, $r-i$, and $i-z$ for the matched spectroscopic catalogue and weak lensing sample on the top row while we show $g-r$ vs $r-i$, $r-i$ vs $g-i$ and $i-z$ vs $r-z$ in the bottom row. The weighted colors of the matched spectroscopic catalogue are a good match to those of the weak lensing sample, although we can see in the middle panel of the bottom row that the tails of the color distributions of the weak lensing sample are not as well approximated. This is due to the fact the matched spectroscopic catalogue only has ∼40 000 galaxies while the weak lensing sample has more than 3 000 000, and hence the tails of the distributions of the weak lensing sample are poorly sampled by the limited amount of objects in the matched spectroscopic catalogue.

We find that 1.6% of the weak lensing sample falls outside the range of colors sampled by our spectroscopic catalogues. It is relatively straightforward to remove these regions, but the results in this work are robust to the inclusion or exclusion of these 1.6% galaxies.

C. Assessing the weighted spectroscopic sample

The weighting procedure assumes that small regions of color-magnitude space (pixels) populated by galaxies in the weak lensing sample are fairly sampled in the matched spectroscopic catalogue. If this is the case, then weighted estimates of performance metrics will be equivalent to those obtained from a complete spectroscopic sample (i.e., one without biases due to a selection function or incompleteness). However, it is possible that some galaxies live in color-magnitude regions where incompleteness could lead to missing populations from the spectroscopic sample. The redshifts of the spectroscopic sample in these regions could then be biased relative to the full sample of DES galaxies that lie in the same regions of color space.

The only sizable sample that we have access to with target selection of comparable depth to DES is the VVDS Deep survey. This sub-survey within VVDS targeted galaxies purely on i-band magnitude at $i < 24$. In order to understand how the incompleteness within this survey corresponds to the color and brightness of the galaxy distribution we break the four-dimensional color-magnitude volume of the weak lensing sample ($g-r$, $r-i$, $i-z$ and i-band magnitude) into cells based on a k-means clustering algorithm [50]. Each cell represents approximately 0.2% of the sample. To each of these k-means cells we assign objects from our weighted spectroscopic and COSMOS photometric redshift samples and objects targeted by the VVDS Deep survey. Within each four-dimensional k-means cell we find the fraction of the VVDS Deep targets that was successfully assigned a high confidence redshift (flag 3, 4, 9, 13, 14 or 19). In Fig. 7 we show the number of VVDS Deep targets and success rate (completeness) in color-color space for three ranges in i-band magnitude. Between them, these magnitude ranges cover the peak of the number counts in the shear catalogue.

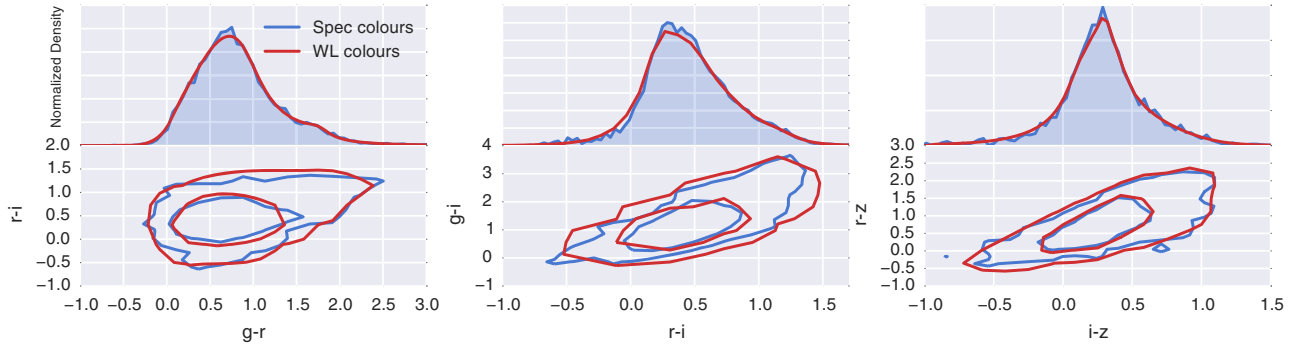

FIG. 6. The color distribution of the weighted matched spectroscopic catalogue is shown in blue relative to the weak lensing sample in red. Top row: One-dimensional histograms of the three colors: $g-r$, $r-i$, and $i-z$. Bottom row: Related two-dimensional comparisons of the color distributions. In general the weighted matched spectroscopic catalogue color distribution matches the weak lensing sample color space well, although in the bottom row we can see that the weighted matched spectroscopic catalogue is unable to match the tails of the weak lensing sample.

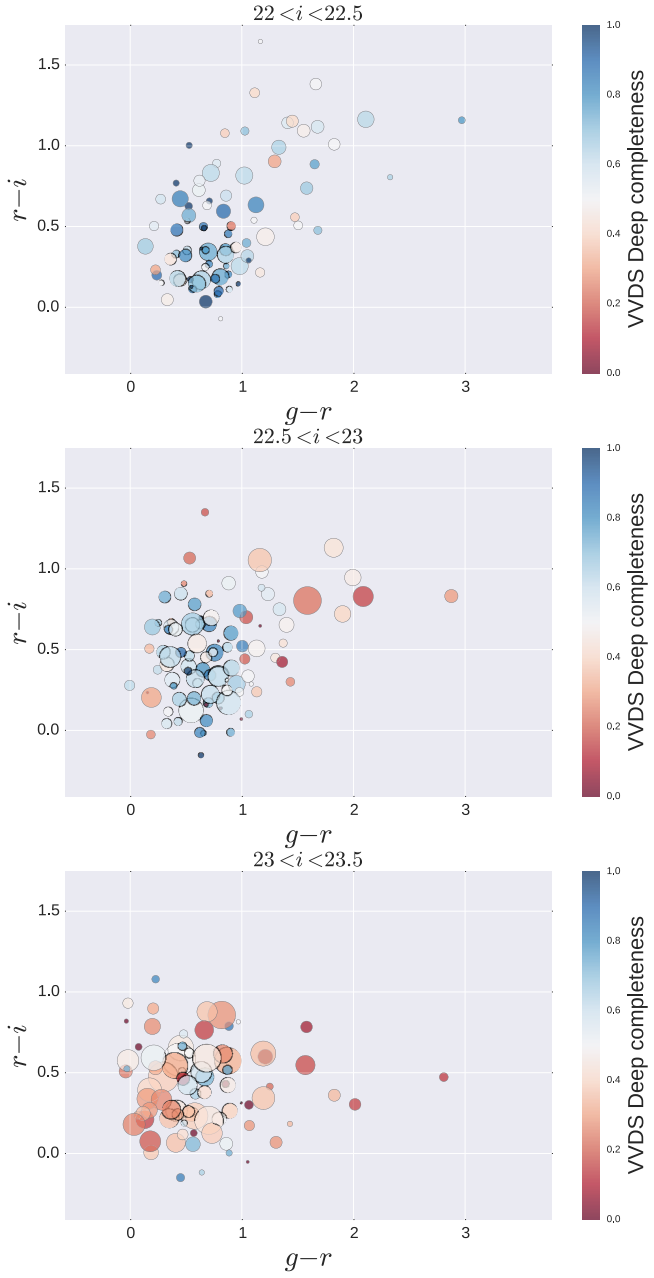

FIG. 7. Spectroscopic completeness of the VVDS Deep sample in $g-r$ vs $r-i$ color space. Each point represents the center of a four-dimensional color-magnitude k-means cell containing a similar number of galaxies from the DES SV NGMIX catalogue. The size of the point represents the number of targeted objects, while the color indicates the fraction that returned a reliable redshift. The three magnitude ranges (as labeled) cover the i-band magnitude range that contains the majority of galaxies in the weak lensing sample; see Fig. 1 for the distribution in the catalogues.

At relatively bright magnitudes ($i < 22.5$) the overall completeness is relatively high, but even here there are typically 20% or more of the targeted galaxies that we do not know the redshifts for. If the incompleteness is due to the clear spectral features of the remaining 20% falling outside of the spectroscopic window then it is easy to imagine that the weighted redshift distribution representing this region of color-magnitude space would be biased. At fainter magnitudes the incompleteness increases, first for the reddest objects, but eventually at $i > 23$, the majority of subsamples are less than 50% complete. We cannot remove weak lensing galaxies in all of the incomplete cells without discarding the majority of our sample. Instead, we try to estimate the likely impact of this incompleteness and in particular whether the uncertainties on the inferred means are consistent with the rest of the uncertainties that we estimate in this work.

In order to estimate the possible impact of incompleteness on the mean redshift of the population we split the color space cells shown in Fig. 7 into regions we term "good" and "bad." The regions are divided at a completeness of 65%, which is the median value of the completeness in the cells. We then compare the mean redshift of the weighted spectroscopic sample to the mean from the photometric redshift catalogue published in Ref. [51] in the COSMOS field, ensuring we use the matched cuts from Sec. II D. Due to the fact that the spectroscopic sample contains many more bright objects than faint ones, only one quarter of the ∼40 000 spectroscopic objects are contained in "bad" cells. We find the difference in the means of the "good" sample is $\delta z = 0.013$, while $\delta z = 0.03$ for the "bad" regions. These errors are comparable to the expected Poisson errors (which alone should be at the level of 0.01) and sample variance (at the level of 0.03), which for a COSMOS-sized survey dominates over Poisson errors for samples with more than 1000 galaxies (see Appendix A of Ref. [11]). For the sample as a whole we therefore do not find evidence for biases in the mean at the level of precision allowed by the samples available.

Later, in Sec. VI, we will see that lensing measurements tend to be dominated by galaxies at higher redshifts. These in turn tend to come from regions with lower levels of completeness. To study this briefly we repeat the comparison between the weighted spectroscopic estimates and the COSMOS samples by first selecting galaxies from the highest redshift bin that we study later ($0.83 < z < 1.3$; see Sec. V). We find differences in means of 0.015 and 0.05 for the good and bad regions respectively. The samples for this study are significantly smaller. The good regions have 624 and 4255 galaxies in the spectroscopic and COSMOS samples, respectively, and the difference in their means can be explained by Poisson errors alone. The bad regions have 1507 and 17 322 galaxies and so the difference in the mean between the spectroscopic and COSMOS determinations cannot be fully explained by Poisson errors alone. However, like the full sample considered above, the difference is similar to that expected from sample variance. We thus conclude that any errors coming from incompleteness for the studies used in this paper are likely to be below the 5% level.

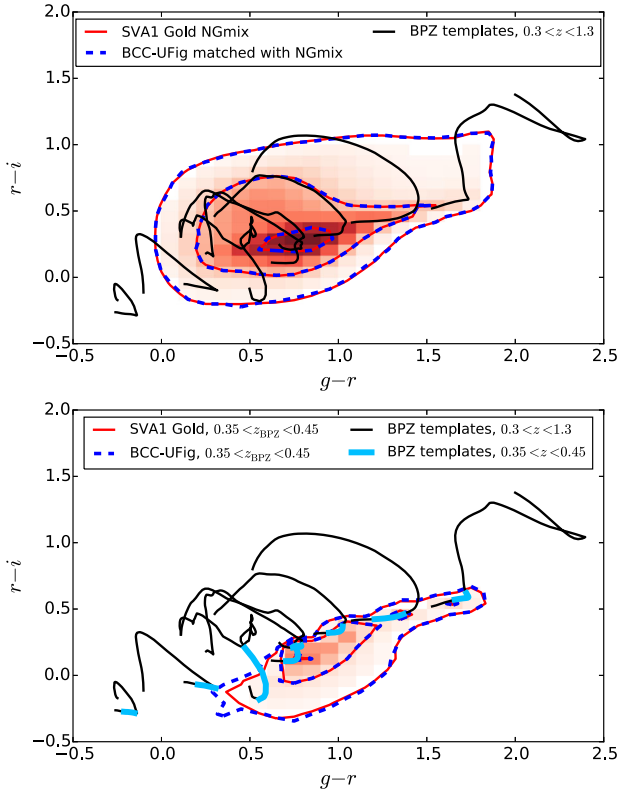

FIG. 8. Upper panel: Color-space distribution of weak lensing sample galaxies and the matched sample taken from BCC-UFig in logarithmic number density intervals (red and blue contours respectively). Over-plotted are the observer-frame colors of redshift-evolved galaxy templates (black lines). Here we show the default set of templates included in the BPZ photometric redshift code, restricted to $0.3 < z < 1.3$ for clarity. Lower panel: The weak lensing and BCC-UFig samples are restricted to objects with BPZ-derived mean redshifts in the range $0.35 < z < 0.45$. The bold light blue sections of the template tracks indicate the same redshift interval for the galaxy models.

D. Biases due to template color coverage

An alternative approach to estimating redshifts empirically based on spectroscopic training samples (e.g., via a ML technique) is to use a set of galaxy templates to fit for galaxy redshifts. By capturing the rest-frame properties of galaxy spectral energy distributions (SEDs), this modeling approach has the advantage that it can be used to interpolate over regions where there are gaps in spectroscopic samples and to extend to higher redshifts. However, as with all modeling approaches there is a risk of introducing model biases if the templates used for the fitting are not fully representative of true galaxies.

In this work we have focused on the set of BPZ templates of Refs. [60,61][1] which like many templates, are built for $z = 0$ galaxies. These do not explicitly account for evolution of the red sequence or changing dust properties at high z. The upper panel of Fig. 8 shows the color space distribution of the weak lensing galaxy sample (red contours) and the matched BCC-UFig sample (blue dashed contours) compared with the observed-frame colors of the BPZ templates redshifted over the range $0.3 < z < 1.3$.

Sharp and strong features in galaxy SEDs, such as the 4000 Å break, create an outer envelope of template colors in certain color-color projections. Of particular importance to this work is where the 4000 Å break transitions between the g and r DECam filters, resulting in extrema in the colors of many templates at $z \sim 0.4$. The effect of this is shown in the lower panel of Fig. 8, where the blue bold sections of the template tracks correspond to the redshift interval $0.35 < z < 0.45$. There is clearly a fairly large region of color-color space, to the bottom-right of the envelope sampled by the template set, for which the closest template will be at $z \sim 0.4$ (in this projection at least). We plot contours for the weak lensing and BCC-UFig samples that lie within this same $0.35 < z < 0.45$ range, showing that indeed the vast majority of galaxies in this region have a redshift solution at $z \sim 0.4$. Previous efforts have in part circumvented this problem, even when using the same template set, by the addition of further photometric bands, in particular the u-band. Expanding the wavelength coverage with additional bands reduces the reliance on single informative colors for redshift determination. In this way, potential bias introduced from template fitting is reduced. For the DES SV data the u-band is not observed, but in Sec. IV we show how we use the BCC-UFig simulations to correct to first order for this effect due to the templates' color coverage.

IV. GLOBAL PHOTO-z BEHAVIOR AND PERFORMANCE

Given the inherent challenges and potential biases in estimating redshifts, we have implemented a number of independent methods for estimating the redshift distribution of the DES SV shear catalogue. Beginning with the global galaxy distribution, we adopt three approaches. The first is an empirical approach based on machine learning methods using spectroscopic training. The second approach is model based and uses a combination of galaxy templates and calibration using the BCC-UFig simulations. Finally we also estimate the galaxy distribution by matching to COSMOS photo-z data. Agreement between the results can give us confidence that possible systematic errors are subdominant, and the level of discrepancy gives an indication of the level of uncertainty that propagates through to later cosmological constraints.

(i) *Empirical spectroscopic*: Several machine learning photo-z methods have been explored within the DES Collaboration, some of which have been previously described in Ref. [26]. In the work that follows we

[1]Multiple template-fitting codes and template sets were used in the preparation of this work, though we present a single choice for brevity.

focus on a subset of these methods, namely ANNZ2, SKYNET and TPZ, which are described in more detail in Appendix C. We note that TPZ and SKYNET do not use the weights in training while ANNZ2 calculates its own weights that it uses in training.

(ii) *Modeling*: For the model-based approach we have implemented the template-based method BPZ. We construct the prior as described in Ref. [60] by fitting to the training sample of the weighted matched spectroscopic catalogue. Using the same prior presented in Ref. [26] has little impact on the results. To calibrate this method we employ a simple first-order correction by applying weak lensing selection cuts to the BCC-UFig catalogues (see Sec. II E) and measuring the offset of the mean redshift between these galaxies and that estimated from the pure BPZ $n(z)$. We find this offset to be 0.050.[2] This offset is applied as a shift to all the BPZ results below, i.e., $n(z) \to n(z - \delta z)$, unless stated otherwise, and is designed to counteract, to leading order, the effect of the peak at $z \sim 0.4$ due the template coverage issues (see Sec. III D) that are present in both the SV data and simulations.

(iii) *Empirical photometric*: The COSMOS field has been observed using DECam and processed through the DES Data Management pipeline to produce coadd images of similar depth to the main SV survey field. Galaxies detected in these images are matched to the [51] photo-z catalogues and then cuts designed to replicate weak lensing selection are applied, as outlined in Sec. II D. Though the photo-z estimates for the COSMOS galaxies are far better than those we can derive from the five DES bands, this approach is limited by sample variance.

For all the results presented in the sections that follow, we retain $0.3 < z_{\mathrm{SkyNet}} < 1.3$ galaxies only. Redshifts of galaxies outside this range are both poorly estimated and have very little impact on the lensing measurements. Galaxies at low redshift have little lensing signal and there are so few at higher redshift that they can be dropped from the analysis. The redshift cuts are made using the SKYNET mean, since we have baselined this method as our default, but results that we present are robust to this choice.

The lower panel of Fig. 9 shows our reconstruction of the $n(z)$ for the DES SV weak lensing sample. The yellow curve comes from the weighted validation set spectra, which is in effect also an estimator of the global distribution. We also show the results of the three machine learning methods, the modeling-based method using BPZ and BCC-UFig and the matched COSMOS results. The vertical lines in the plot show the means of the distributions, which are also listed in Table II. We focus on the mean since it is well known that uncertainty in the mean is the first-order cause of systematic errors in weak lensing [7]. Later, in Sec. VI, we will propagate the full errors through to weak lensing

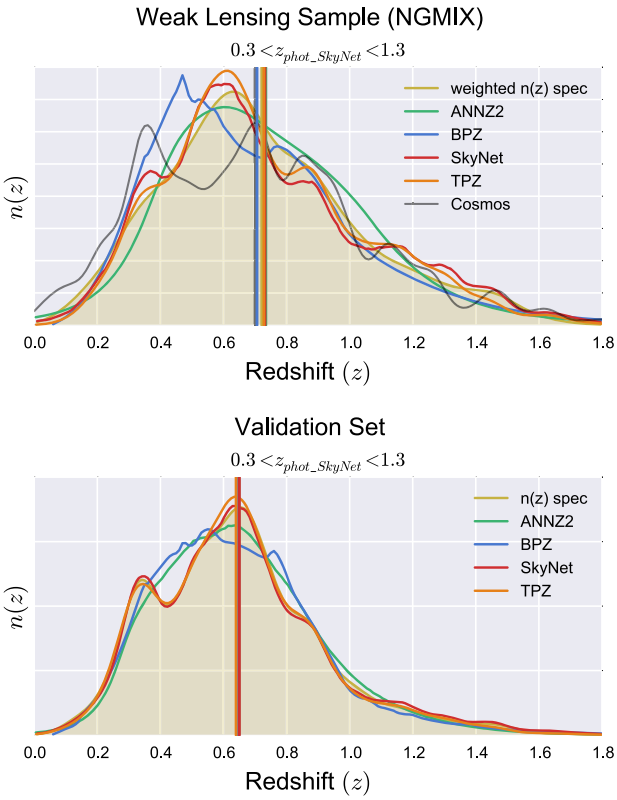

FIG. 9. The full redshift distribution $n(z)$ (i.e. stacked probability density function) for the validation sample ($0.3 < z < 1.3$). Upper panel: The kernel density estimate of the full unweighted validation sample compared to the four photo-z methods. Lower panel: The same, but including the weighting from Sec. III B and matched COSMOS photometric redshifts from Ref. [51]. The vertical lines in the plots are the mean values of the distributions.

TABLE II. The left column contains the estimates of the mean of the redshift distribution of the NGMIX sample of the four photo-z methods and also the mean of the weighted spectroscopic sample which is itself an estimate of the mean of the NGMIX sample. The right column contains the mean of the unweighted validation set with the four photo-z methods and the mean from the spectra.

Spectra	DES SV -WL sample 0.72 (weighted)	Validation sample 0.64
ANNZ2	0.73	0.65
SKYNET	0.73	0.65
TPZ	0.73	0.64
BPZ	0.71	0.64
Matched COSMOS	0.70	–

[2]Though the BCC-UFig sample is color matched to the weak lensing sample after performing the initial weak lensing cuts, this does not influence the correction. If we do not color match the BCC-UFig sample to the weak lensing sample, we find an offset of 0.049.

statistics and σ_8. We see that all of our estimates of the global distribution of galaxies give comparable results and we estimate the mean to be 0.72 with a precision better than 0.02. As a further test, we also show results when we apply the same procedure to the unweighted validation sample. Here we take the spectroscopic sample to be a truth catalogue and we can see again that our methods are able to find the mean of this distribution to a precision better than 0.01. The corresponding means for these results are also shown in Table II. In the top panel of Fig. 9 we see that redshift distributions of TPZ and SKYNET seem to have radial features that are also present in the spectroscopic

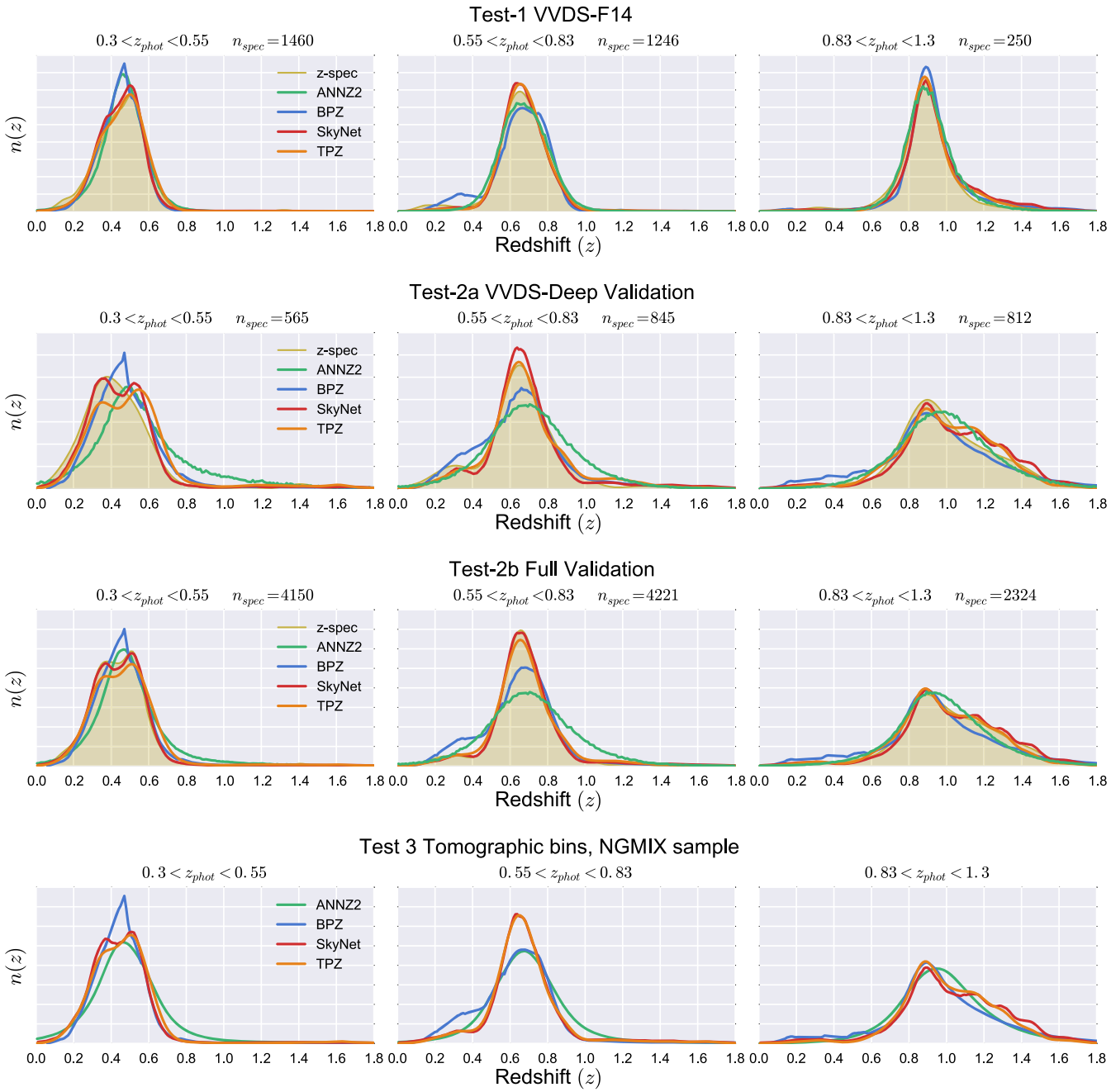

FIG. 10. Each row of panels shows the weighted spectroscopic redshift distributions (shaded area) of the objects in each tomographic bin as selected by the mean of SKYNET compared to estimates of the redshift distribution of the four methods used in this work. Top row: The spectra used in this test comes from VVDS-F14, an independent sample not used for training. Second row: The spectra used in this test are a 30% subset of VVDS-Deep used as part of the validation sample. Third row: The spectra used in this test are a 30% subset of the matched spectroscopic catalogue used for validation. Bottom row: The redshift distribution in the tomographic bins for the NGMIX sample.

training set; this is likely due to the fact that TPZ and SKYNET are given too much freedom to model the probability density function (PDF), and this has no effect on the cosmological inference in SVA-1.

V. TOMOGRAPHIC PHOTO-Z PERFORMANCE

In the previous section, we discussed the global characteristics of the estimated $n(z)$. In the cosmological analysis of Ref. [20], we have presented a conservative analysis of the two-point cosmic shear constraints on cosmology by marginalizing over a large array of nuisance parameters related to known or suspected systematics. Particularly in the case of intrinsic alignment, doing so severely degrades the constraining power of a nontomographic analysis. Thus we must also characterize how well the four photo-z methods are able to reconstruct the redshift distribution of individual tomographic bins—in this case, three bins selected that match those used in Refs. [20,62]. These are designed to contain approximately equal lensing weight in the larger NGMIX shear catalogue. The bin boundaries are set by cuts on the SKYNET mean redshifts at [0.3, 0.55, 0.83, 1.3]. We choose to keep the galaxies in each bin fixed according to the cosmology analysis of Ref. [20].

In this section we look at the photo-z performance in these three tomographic bins. This is done through a series of tests, comparing the reconstruction of $n(z)$ (and in particular the value of the mean redshift) in three spectroscopic galaxy samples and the NGMIX catalogue:

(i) *Test 1*: An independent sample of spectroscopic galaxies in the VVDS-F14 field, which were not used in training or validation and located in a distinct part of the sky separate from the training and validation fields. The radial structure in the independent sample is thus different from what the machine learning methods trained on.

(ii) *Test 2a*: A deeper spectroscopic sample of 30% of the galaxies in the VVDS-Deep field, which matches better to the depth of DES SV photometry, but which is also part of the validation sample and thus not fully independent.

(iii) *Test 2b*: The full validation sample—30% of the matched spectroscopic sample set—excluding galaxies in the VVDS-F14 field.

(iv) *Test 3*: Comparison of the redshift estimates of the four photo-z methods for the full DES SV NGMIX catalogue.

Once again, we use SKYNET as the fiducial photo-z result, and so for consistency all objects in this section are assigned a bin based on the mean of the SKYNET $p(z)$. In Appendix B, we show results where each code assigns a bin to each galaxy based on their own z mean. Figure 10 shows the results in the tomographic bins of Tests 1, 2a and 2b for each of the photo-z algorithms we consider as labeled. Overall we see that all the methods produce consistent results. Since we do not have a perfectly representative spectroscopic sample for the galaxy population for the full NGMIX catalogue, we only compare the relative agreement of the photo-z methods in the bottom panel of Fig. 10. The bin with the highest cosmological information content for tomographic lensing is the highest redshift bin. It is therefore reassuring that visually the different methods give consistent results. Table III shows the mean offsets of the results shown in the top three panels of Fig. 10. Table IV shows the estimates of the mean in the tomographic bins of the NGMIX sample by the photo-z codes and the estimate of the weighted spectroscopic sample. We see from the results for Tests 2b and 3, which are the closest to our weak lensing samples, that the relative bias of the means are broadly consistent with Gaussian scatter of width 0.05.

TABLE III. The bias $(\langle z_{\rm phot}\rangle - \langle z_{\rm spec}\rangle)$ between the photometric redshift estimates and the true spectroscopic distribution in Test 1 ("independent"), Test 2a (VVDS-Deep) and Test 2b (full validation set).

	z range	ANNZ2	BPZ	SKYNET	TPZ
Test 1	0.30–0.55	0.014	0.001	0.003	0.008
	0.55–0.83	0.019	−0.002	0.017	0.017
	0.83–1.30	0.033	0.057	0.063	0.039
Test 2a	0.30–0.55	0.139	0.072	0.027	0.079
	0.55–0.83	0.069	0.027	0.034	0.042
	0.83–1.30	0.002	−0.026	0.044	0.016
Test 2b	0.30–0.55	0.064	0.032	0.012	0.033
	0.55–0.83	0.027	−0.010	0.013	0.010
	0.83–1.30	−0.030	−0.045	0.022	−0.016

TABLE IV. The estimated mean of the three tomographic bins in the NGMIX sample of the four photo-z methods and the estimate of the weighted spectroscopic sample.

z range	Spec (weighted)	ANNZ2	BPZ	SKYNET	TPZ
0.30–0.55	0.45	0.49	0.46	0.45	0.46
0.55–0.83	0.67	0.69	0.64	0.67	0.67
0.83–1.30	1.00	0.98	0.97	1.02	1.01

VI. IMPLICATIONS FOR WEAK LENSING

The mapping of traditional photo-z metrics to actual impacts on the weak lensing measurements and cosmological parameter constraints is nontrivial, and the resulting bias can be difficult to capture using simple metrics. In this section we explore the impact of photo-z uncertainty by propagating the errors through the two-point correlation function and to the cosmological parameter σ_8 and to measurements of $\Sigma_{\rm crit}^{-1}$.

A. Photo-z impact on two-point cosmic shear analysis

The photo-z $n(z)$ impacts the predicted correlation function (and thus constraints on cosmological parameters)

through the lensing efficiency when modeling the convergence power spectrum $C(\ell)$. The tomographic correlation function $\xi_{+/-}$ is related to $C(\ell)$ through the zeroth (fourth)-order Bessel function of the first kind by

$$\xi_{+/-,ij}(\theta) = \frac{1}{2\pi}\int d\ell\ell C_{ij}(\ell) J_{0/4}(\ell\theta), \tag{1}$$

where $(i,j) \in (1,2,3)$ represent the redshift bins in the auto- or cross-correlation. $C_{ij}(\ell)$ is then defined as

$$C_{ij}(\ell) = \int_0^{\chi_H} d\chi \frac{W_i(\chi)W_j(\chi)}{\chi^2} P_\delta\left(\frac{\ell}{\chi},\chi\right), \tag{2}$$

for comoving distance χ, horizon distance χ_H, matter power spectrum P_δ, and lensing efficiency, given in a flat universe as

$$W_i(\chi_l) = \frac{3H_0^2\Omega_m}{2c^2}(1+z_l)\chi_l \int_{\chi_l}^{\chi_H} d\chi_s n(\chi_s)\frac{\chi_s-\chi_l}{\chi_s}. \tag{3}$$

The redshift distribution of galaxies is normalized such that $\int n_i(\chi)d\chi = 1$, H_0 is the Hubble parameter, and Ω_m is the matter density parameter at $z = 0$.

The predicted $\xi_{+/-}$ (both tomographic and nontomographic) are calculated over the θ range and tomographic binning is used for the measurements in Ref. [62] for each photo-z estimate and the weighted matched spectroscopic sample. We then use these predicted correlation functions with the covariance matrix from Ref. [62] to propagate the differences between photo-z estimates through to constraints on σ_8 (with all other parameters fixed). The "truth" (or measurement of $\xi_{+/-}$ with no systematic uncertainties) is taken to be either the fiducial SKYNET prediction in Sec. VI A 1 or the weighted matched spectroscopic sample in Sec. VI A 2, while each photo-z estimate's predicted $\xi_{+/-}$ is taken to be the assumed theory in turn when constraining σ_8. The final results of this comparison for the four photo-z estimates presented in this work are shown in Figs. 11–13.

1. Comparison of photo-z estimates for the DES SV shear catalogue

For the full photometric galaxy sample contained within the shear catalogue, we have no estimate for the true value of the $n(z)$ to compare to and so instead compare to the fiducial SKYNET prediction as a relative point of reference. We can therefore only compare the relative agreement between the photo-z codes shown for the NGMIX catalogue in Fig. 11.

In the left panel, the relative agreement in the magnitude of ξ_+ is shown, averaged over θ.[3] The left set of points show the nontomographic ξ_+, while the middle and right sets of points show the three auto- and cross-correlations, respectively. The grey bands show the 1σ error on the magnitude of the measured ξ_+ for each correlation function, using the covariance calculated in Ref. [62]. The relative agreement in ξ_+ between the machine learning methods is very good in correlations with the highest tomographic bin ("33," "23," and "13"). This increases significantly for correlations with the lower tomographic bins ("11," "22," and "12"), though the nontomographic case also has good agreement on the order of 5%. BPZ tends to disagree with the machine learning methods, typically at the 5–10% level.

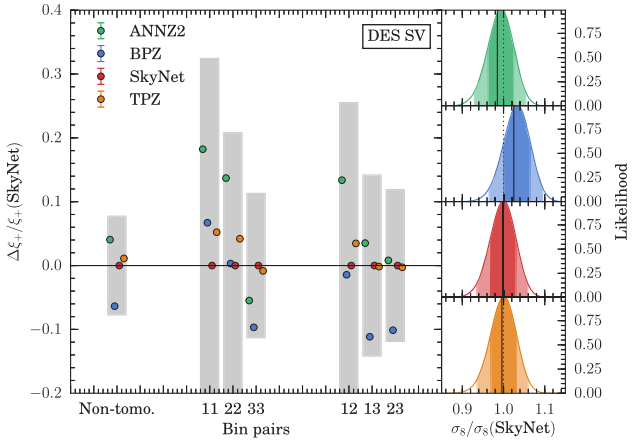

FIG. 11. A comparison of the relative agreement of the $n(z)$ estimates for ANNZ2, BPZ, SKYNET, and TPZ for the NGMIX shear catalogue. Left panel: The relative magnitude of the correlation function compared to the SKYNET $n(z)$ prediction is shown for the nontomographic ξ_+, the three autocorrelations, and the three cross-correlations. The grey band is the actual variance in the magnitude of ξ_+ measured from SV data. Right panels: The corresponding constraints on σ_8, with fiducial SKYNET results normalized to one (vertical dotted black line). The likelihood histograms, color-coded to match the ξ_+ points on the left, are shown for each tomographic constraint. The peak of the likelihood histogram for the nontomographic constraint is given by the vertical black line for comparison. The vertical ordering is the same as the legend in the left panel.

The right panels of Fig. 11 show the corresponding constraints on σ_8. The SKYNET prediction is normalized to one (vertical dotted black line). The likelihood histogram, colored to match the points in the left panel for each photo-z code, is shown for the full tomographic constraint, while the vertical solid black line gives the peak of the likelihood histogram for the nontomographic constraint. The bias in constraints on σ_8 between the machine learning photo-z methods is very small despite low-z differences in the correlation function, with agreement at much better than the 1σ level. BPZ has a relative bias of about 1σ, by comparison, which corresponds to about 3% in σ_8.

For completeness, we have also repeated the above analyses and those in Sec. VI A 2 on the IM3SHAPE $n(z)$ with the same redshift boundaries matching those derived

[3]The major results are unchanged when instead considering specific values of θ.

for NGMIX and again for tomographic bins derived for IM3SHAPE, and find in all cases that the major conclusions and resulting differences across photo-z methods are consistent between analyses of the two catalogues at the level of accuracy we require for SV analysis.

2. Null tests relative to matched spectroscopic samples

One difficulty with the results in Sec. VI A 1 is that we have no way of determining what the true $n(z)$ is, and thus can only compare relative agreement between photo-z methods. We can, however, create an experiment in which the $n(z)$ is known to be exactly that of our weighted independent spectroscopic sample (Test 1). We then repeat the analysis from Sec. VI A 1 for this test as an additional way of characterizing systematic photo-z uncertainties. Though there are only 2956 galaxies in the independent spectroscopic sample within our $0.3 < z < 1.3$ boundaries, we assume the estimated $n(z)$ from each code and the spec-z distribution instead represents a sample with the same number of objects as the NGMIX catalogue. These redshift distributions (see top panel Fig. 10) are used to measure the relative difference in $\xi_{+/-}$ compared to the spectroscopic prediction as in Sec. VI A 1. We also calculate error bars on the points, which represent the 1σ error in the difference from bootstrapping the $n(z)$ of the sample. Since we are comparing the matched photometric and spectroscopic $n(z)$ distributions for the same galaxies contained within the VVDS-F14 field, there is no sample variance contribution to these error bars. However, since it is a small field separate from the DES SV SPT-E region, any extrapolation of the bias to the full DES SV shear catalogue could still be over- or under-estimated.

We show the results of this analysis in Fig. 12. The bias in ξ_+ relative to the spectroscopic prediction for the three machine learning codes (ANNZ2, SKYNET, and TPZ) is shown in the left panel. It is in good agreement and consistent across the correlations at about 5–10% larger than the spectroscopic prediction. This is consistent with the machine learning codes producing too wide $p(z)$ or over-estimated high-z tails, both of which can bias ξ_+ high. The empirically corrected BPZ photo-z estimates perform similarly, with a maximum bias in ξ_+ of 10% in the highest redshift autocorrelation.

The right panels of Fig. 12 show the corresponding constraints on σ_8. The weighted spectroscopic prediction is normalized to one (vertical dotted black line) and the vertical grey band is the 1σ bootstrap error corresponding to the error bars on the ξ_+ points. Note, however, that discussion of deviations in σ_8 will refer primarily to the marginalized constraints unless specifically referring to the bootstrap error. The tomographic and nontomographic constraints agree well. All four photo-z estimates are biased slightly low by just less than 1σ. It is important to note that due to the small sample size in the independent spectroscopic test sample, the 1σ bootstrap error in σ_8 just due to

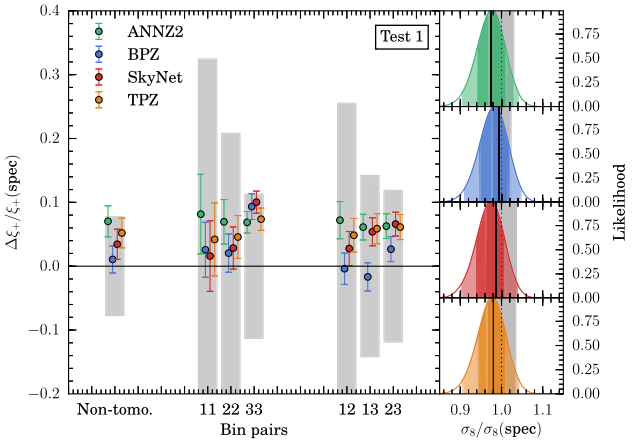

FIG. 12. A comparison of the relative agreement of the $n(z)$ estimates for ANNZ2, BPZ, SKYNET, and TPZ to the weighted independent spectroscopic galaxy sample. Left panel: The relative magnitude of the correlation function compared to the spectroscopic $n(z)$ prediction is shown for the nontomographic ξ_+, the three autocorrelations (11, 12, 33 bin pairs), and the three cross-correlations (12, 13, 23 bin pairs). The grey band is the actual variance in the magnitude of ξ_+ measured from SV data. Error bars on the points are the 1σ error on the difference of ξ_+ obtained from bootstrapping the $n(z)$ of the spectroscopic sample. Right panels: The corresponding constraints on σ_8, normalized to one (vertical dotted black line). The likelihood histograms, color-coded to match the ξ_+ points on the left, are shown for each tomographic constraint. The peak of the likelihood histogram for the nontomographic constraint is given by the vertical black line for comparison. The vertical grey band is the corresponding 1σ bootstrap error in σ_8.

sample variance in the independent spectroscopic sample is of the same order as the 1σ constraints on σ_8 in DES SV for some methods. Overall, we find a level of systematic bias from this test in σ_8 of 1–3%.

We note that in the case where we use the SVA-1 covariance to model the uncertainty on ξ_+, the error on the correlation function in the tomographic case is due in large part to the contribution of uncorrelated shape noise in the tomographic bins. When combining the tomographic cross-correlations, this leads to a much better constraint on σ_8 than the amplitude of any individual cross-correlation. In the case of the bootstrap error on the $n(z)$ (and thus ξ_+, σ_8), no shape noise is present and the variation is only due to the change in $n(z)$ from the bootstrap samples. Here the correlation between the $n(z)$ changes in the tomographic bins is very high and thus the gain of information is minimal when combining the various cross-correlations. This explains the fact that the error on σ_8 in the left panels of Fig. 12 due to the bootstrap error is only slightly smaller than the error of the SVA-1 ξ_+ scenario (right panel).

We can further diagnose the performance of the photo-z codes' estimates of the $n(z)$ by considering the Bayes factor

TABLE V. Values of $\ln K$ for the Bayes factor $K = Pr(D|p_1)/Pr(D|p_2)$ are shown for each photo-z estimate (p_1, rows) compared to another (p_2, columns) when constraining the value of σ_8 [all other cosmology is kept fixed, varying only the estimates of $n(z)$ between p_1, p_2, and D]. The values for tomographic (nontomographic) analyses in Figs. 12 and 13, and the right panel of Fig. 14 are given. The Bayes factor gives an indication of how much more supported one photo-z estimate (p_1) is than another (p_2) by the data D, in this case the predicted correlation function built from the weighted spectroscopic estimate of $n(z)$. A value $\ln K > 1$ generally indicates that p_1 is more strongly supported as the true photo-z estimate.

		ANNZ2	BPZ	SKYNET	TPZ
Test 1	ANNZ2	–	0.36 (−0.04)	0.22 (−0.02)	0.03 (−0.01)
	BPZ	−0.36 (0.04)	–	−0.13 (0.01)	−0.33 (0.03)
	SKYNET	−0.22 (0.02)	0.13 (−0.01)	–	−0.2 (0.01)
	TPZ	−0.03 (0.01)	0.33 (−0.03)	0.2 (−0.01)	–
Test 2a	ANNZ2	–	−3.94 (−0.1)	−7.02 (−0.04)	−5.2 (−0.06)
	BPZ	3.94 (0.1)	–	−3.08 (0.07)	−1.26 (0.04)
	SKYNET	7.02 (0.04)	3.08 (−0.07)	–	1.82 (−0.02)
	TPZ	5.2 (0.06)	1.26 (−0.04)	−1.82 (0.02)	–
Test 2a Corrected	ANNZ2	–	−0.08 (−0.02)	−0.08 (−0.02)	−0.04 (−0.01)
	BPZ	0.08 (0.02)	–	0.0 (−0.0)	0.04 (0.01)
	SKYNET	0.08 (0.02)	−0.0 (0.0)	–	0.04 (0.01)
	TPZ	0.04 (0.01)	−0.04 (−0.01)	−0.04 (−0.01)	–

$$K = \frac{Pr(D|p_1)}{Pr(D|p_2)}, \tag{4}$$

where Pr is the posterior probability of the model p_i due to some photo-z estimate in the σ_8 constraints of Fig. 12. In this analysis, D refers to the predicted $\xi_{+/-}$ for the weighted matched spectroscopic samples, and Pr is the integrated posterior likelihood. The Bayes factor can be used to compare how well supported by the data two models are. A value $\ln K > 1$ supports p_1 over p_2, with p_1 being substantially supported when $\ln K > 3$. The Bayes factor is given for each combination of photo-z estimates in Table V. The Bayes factors from the tomographic analysis are given first, with the nontomographic Bayes factors shown in parentheses for comparison. We find that there is no significant preference for one photo-z code over another for the independent sample (Test 1), though there is some evidence that BPZ does slightly worse and ANNZ2 slightly better. This distinction is lost, however, for the nontomographic analysis, which is unable to differentiate the photo-z estimates.

We also want to compare the photo-z performance of the four codes for a set of spectroscopic redshifts that better match the depth of the DES SV data. Figure 13 instead compares the correlation function and σ_8 constraints for the photo-z estimates of galaxies in the weighted "deep" spectroscopic sample of Test 2a. The predicted $n(z)$ for these galaxies is shown in the third panel in Fig. 10. All four codes perform more poorly for this "deep" sample compared to the analysis of Test 1 in Fig. 12, with a greater spread in the magnitude of the predicted ξ_+ relative to the spectroscopic prediction. SKYNET is the most stable across tomographic bins, with a spread in bias values limited to around 5%. The other codes scatter to a much wider range of values. For the lower bins in particular, there is significant bias in ξ_+. The larger scatter in Test 2a as compared to Test 1 is due to the fact that the photo-z's perform significantly worse for the fainter sample.

The corresponding σ_8 constraints are driven by information in the highest redshift bin, however, and have a more reasonable bias about the weighted spectroscopic prediction. The four photo-z estimates still agree with the matched spectroscopic prediction for the VVDS-Deep sample within 1σ, except for ANNZ2, which is biased at the 2–3σ level. The large range of bias between the lowest and highest redshift bins also produces a nearly 1σ tension

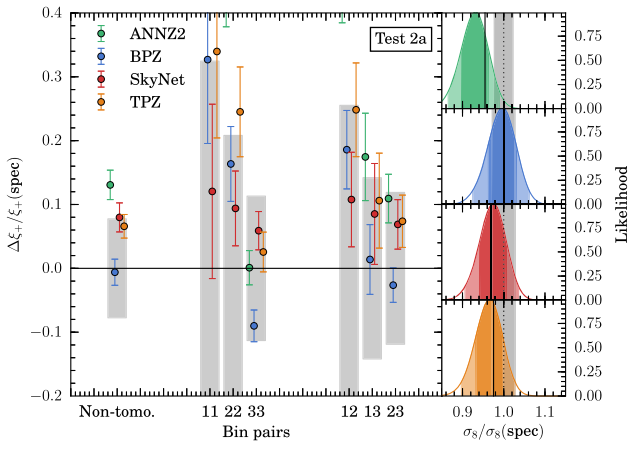

FIG. 13. A comparison of the relative agreement of the $n(z)$ estimates for ANNZ2, BPZ, SKYNET, and TPZ to the weighted "deep" spectroscopic galaxy sample, showing the same information as described in Fig. 12.

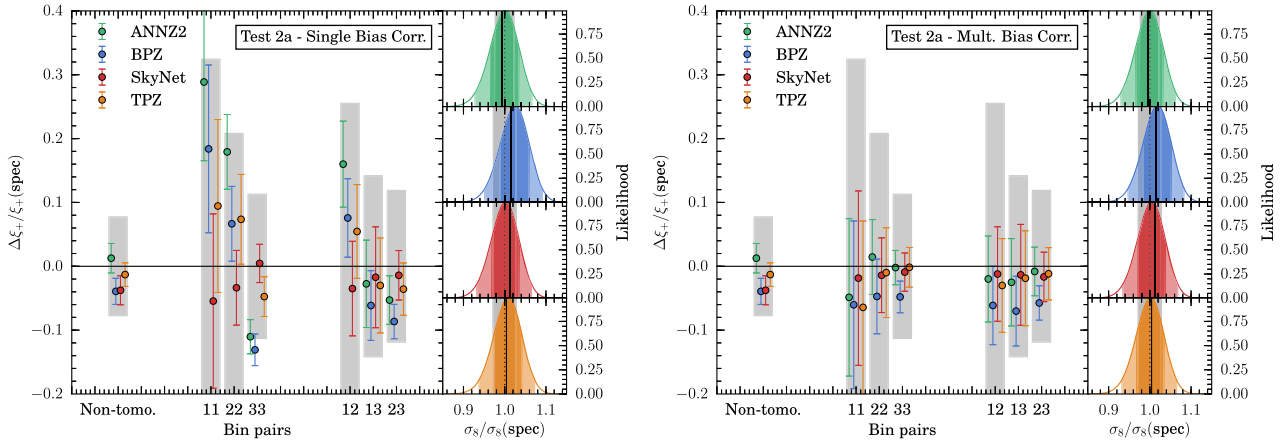

FIG. 14. The effect on Fig. 13 of applying a bias correction to the mean of the $n(z)$ of each photo-z estimate by comparison to the true spectroscopic $n(z)$. The left side fixes a single bias parameter for the three tomographic bins, while the right side allows a different bias parameter for each bin. Each side shows the same information as described in Fig. 12.

between the tomographic and nontomographic constraints for ANNZ2. This bias is not explained as an artifact of selecting the binning of galaxies based on SKYNET, as seen in Appendix B. We present the associated Bayes factor values in Table V, where SKYNET is significantly favored over the other three codes. There is again no distinction between the codes, however, in the nontomographic analysis from the Bayes factor. Overall, we find a maximum level of systematic bias from this test in σ_8 of 7% for ANNZ2, though the bias in the other methods is similar to the level found in Test 1.

B. Validation of priors for photo-z bias parameters

To first order, we can correct for the systematic redshift biases shown in Sec. VI A with the approximation $n_i(z) \to n_i(z - \delta z_i)$ where δz_i is the bias on the mean redshift of the source galaxies in the appropriate tomographic bin. In the cosmology analysis of Ref. [20]. we adopted a Gaussian prior of width 0.05 on the allowed bias values based on comparisons of the four photo-z methods' estimates of the $n(z)$ discussed in Secs. IV and V. This is shown explicitly in Fig. 14, where we compare the impact such a correction scheme has on ξ_+ and σ_8. The bias parameters by which the $n(z)$ are shifted are not marginalized over here, but instead are taken from Table III for Test 2a, since we can directly calculate the bias.

We find that a single mean redshift bias parameter is sufficient to resolve the bias in σ_8 for all four codes. Taking into consideration the 1σ bootstrap error in the ξ_+ ratio, all the tomographic correlations are consistent with zero remaining bias in ξ_+ for SKYNET, and the other photo-z estimates are also greatly improved relative to the spectroscopic prediction. Relaxing this to a bias parameter for each redshift bin does not further significantly improve the bias in σ_8, but it does have a large impact on the agreement in ξ_+, which could have an impact on other parameter constraints. All tomographic points are now consistent with zero for the machine learning methods. This is confirmed in the Bayes factor, shown in Table V for the three-parameter case. All values of K are consistent with the four corrected photo-z estimates being equally likely to be true.

We thus employ a Gaussian prior on the photo-z bias of width $\delta z_i = 0.05$, centered at zero, separately for each of the tomographic bins in the fiducial cosmology analysis of Ref. [20]. We also explore the effect of propagating a nonzero center for the prior in the analysis discussed in that paper, and find no significant differences to the cosmology results.

C. Photo-z impact on other lensing analyses

In general, the main impact of photo-z uncertainties in weak lensing measurements enters through the impact on the critical surface density $\Sigma_{\rm crit}$. This quantity captures the information on distance ratios in lens-source pairs that lensing is sensitive to, namely

$$\Sigma_{\rm crit}^{-1} = \frac{4\pi G}{c^2}\frac{D_{ls}D_l}{D_s}, \tag{5}$$

where D_l is the angular diameter distance to the lens, D_s is the distance to the source, and D_{ls} is the distance between the lens and source. Calculating $\Sigma_{\rm crit}^{-1}$ uses the individual $p(z)$ for each galaxy, which is a different test of the photo-z quality than the bulk summation into large tomographic bins for cosmic shear analysis. It is also possible to directly calculate this quantity for a relatively small sample of galaxies, unlike the correlation function, allowing us to directly compare the photo-z methods' predictions for this quantity with the weighted matched spectroscopic prediction.

To explore this, we compare the impact of the different redshift estimates on the calculation of $\langle\Sigma_{\rm crit}^{-1}\rangle$ as a function of lens redshift. This directly probes the impact of photo-z bias in measurements of $\Delta\Sigma$ in cluster and galaxy-galaxy lensing, and is relevant for other tangential shear measurements where one can distinguish between a population with significantly better photo-z estimates than a source sample in the larger shear catalogue. We will assume that lens galaxies have negligible redshift error relative to the source catalogue and thus have no impact on the calculation of $\langle\Sigma_{\rm crit}^{-1}\rangle(z_{\rm lens})$ for the purpose of evaluating the redshift estimates presented in this paper. We follow the same process as described above, repeating this analysis for the deep matched spectroscopic sample (Test 2a) and for the full DES SV NGMIX shear catalogue.

For each galaxy sample and photo-z estimate, we evaluate the weighted mean inverse $\Sigma_{\rm crit}$ as a function of lens redshift

$$\langle\Sigma_{\rm crit}^{-1}\rangle(z_{\rm lens}) = \sum_i p(z_{{\rm source},i})\Sigma_{\rm crit}^{-1}(z_{\rm lens}, z_{({\rm source},i)}), \qquad (6)$$

where $\sum p(z) = 1$. For the spectroscopic test, $\langle\Sigma_{\rm crit}^{-1}\rangle_{\rm spec}$ is simply evaluated at the spectroscopic redshift with no probability distribution. We use three source redshift bins $0.5 < z_{\rm source} < 1.3$, $0.7 < z_{\rm source} < 1.3$, and $0.9 < z_{\rm source} < 1.3$, as well as the nontomographic range from the two-point analysis, $0.3 < z_{\rm source} < 1.3$. We calculate $\langle\Sigma_{\rm crit}^{-1}\rangle$ over the lens redshift range $0.1 < z_{\rm lens} < 0.9$, which brackets the redshift limits of the lenses in the red-sequence Matched-filter Galaxy Catalog, described in Ref. [63], which selects red-sequence galaxies. The catalogue used here in what follows is limited to luminosity $L > L_*$, which results in approximately 30 000 lenses. To calculate statistical errors for the figures, we use three lens redshift bins: $0.2 < z_{\rm lens} < 0.4$, $0.4 < z_{\rm lens} < 0.6$, and $0.6 < z_{\rm lens} < 0.8$.

Figure 15 shows the resulting $\Delta\langle\Sigma_{\rm crit}^{-1}\rangle/\langle\Sigma_{\rm crit}^{-1}\rangle_{\rm spec}$ for Test 2a. We find good agreement between the photometric estimates and the matched spectroscopic redshifts. SKYNET and TPZ have biases that are nearly consistent with zero in all bins and lens redshifts, reaching levels comparable to the bootstrap errors over the spectroscopic sample at high redshift. The worst performing method, BPZ, has a bias that reaches only 15% at the high-redshift limit. For comparison, we include the statistical error on the magnitude of the tangential shear signal calculated via jackknife of the lens sample over the DES SV footprint. The weighted tangential shear $\gamma_t(\theta)$ enters into the calculation of $\Delta\Sigma$ linearly with $\Sigma_{\rm crit}^{-1}$. Except for BPZ, the bias for all methods is typically much less than this statistical error. We exclude Test 1 due to there being insufficient galaxies in the higher redshift bins to produce a ratio that is not dominated by noise, but have verified that in the lowest redshift bin, for example, there is negligible bias consistent with that shown in Fig. 15 for Test 2a.

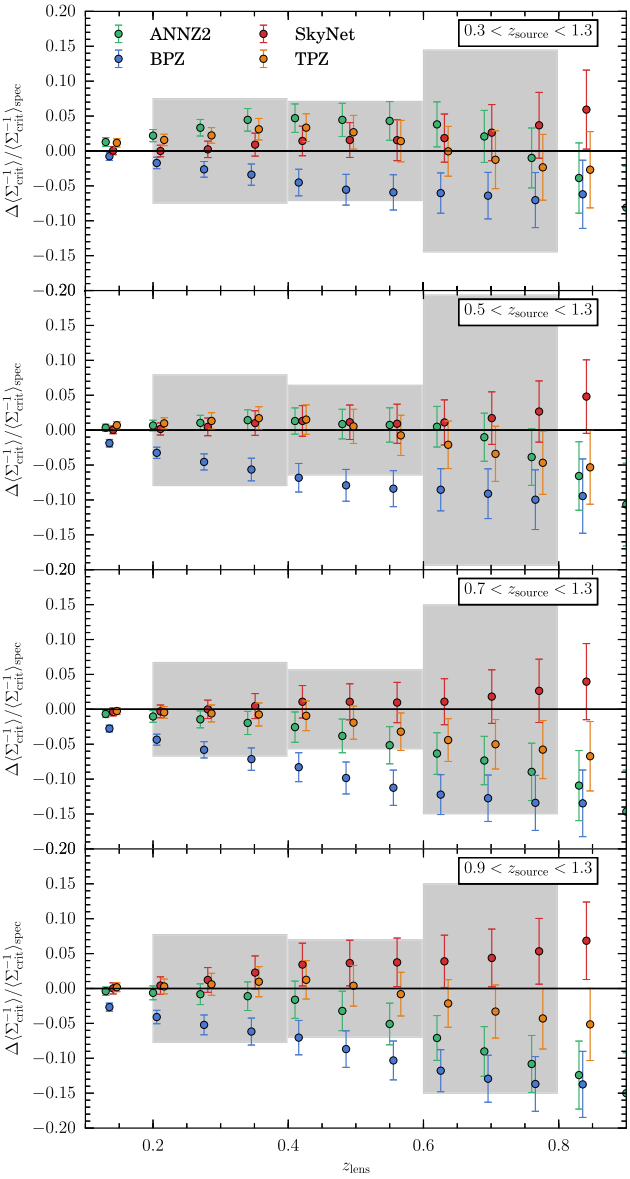

FIG. 15. The fractional difference in the $\langle\Sigma_{\rm crit}^{-1}\rangle$ between the photo-z estimates and the deep Test 2a spectroscopic prediction is shown as a function of lens redshift for four source redshift bins. Grey bands show the 1σ statistical error in the measurement of the tangential shear signal for the three lens bins indicated by the width of the bands.

We repeat the same analysis for the full DES SV NGMIX shear catalogue in Fig. 16. The left panel shows $\langle\Sigma_{\rm crit}^{-1}\rangle$ as a function of lens redshift for each photo-z estimate, which agree well with each other. The differences are quantified in the right panels for each source redshift bin, where the fractional difference from the mean is shown. The spread in relative differences between the codes is within 5% of that seen for the deep Test 2b in Fig. 15, which suggests that the bias shown in Fig. 15 is a good estimate of that expected in DES SV measurements of $\langle\Sigma_{\rm crit}^{-1}\rangle$.

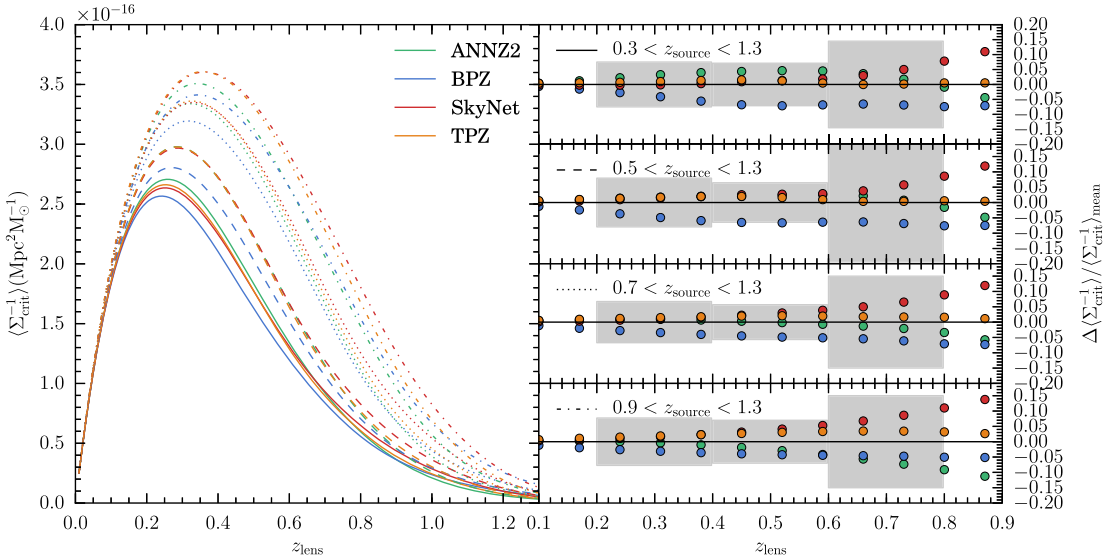

FIG. 16. Left: $\langle\Sigma_{\rm crit}^{-1}\rangle$ for the full NGMIX shear catalogue is shown as a function of lens redshift for four source redshift bins: $0.3 < z_{\rm source} < 1.3$ (solid), $0.5 < z_{\rm source} < 1.3$ (dashed), $0.7 < z_{\rm source} < 1.3$ (dotted), and $0.9 < z_{\rm source} < 1.3$ (dash-dotted). Right: The fractional difference in the $\langle\Sigma_{\rm crit}^{-1}\rangle$ between the photo-z estimates relative to the mean. Grey bands show the 1σ statistical error in the measurement of the tangential shear signal for the three lens bins indicated by the width of the bands.

VII. CONCLUSIONS

The Dark Energy Survey aims over five years of observations to combine the measurements of shapes and redshifts for hundreds of millions of galaxies to constrain cosmological parameters and to study the evolution and structure of dark energy and dark matter. The determination of accurate redshift distributions for these galaxies is one of the primary challenges for DES and for future weak lensing surveys, and may become the dominant systematic limitation in pursuing cosmology through precision weak lensing measurements. We have presented in this work an analysis of the resulting redshift distributions of galaxies with shape measurements from the pre-survey Science Verification data for DES, and identified key challenges and obstacles in the pursuit of producing accurate redshift distributions for the main DES survey data releases at the level required to support ongoing DES weak lensing science.

We have compiled a set of more than 46 000 spectroscopic galaxies, which are matched in image depth and weighted to ensure even sampling of the weak lensing sample. These galaxies are split into training and validation samples, as well as an independent validation sample and a deep validation sample, the latter of which overlaps with the primary validation sample. The independent sample is taken from a separate spectroscopic field (VVDS-F14), while the deep sample is closer to the DES SV magnitude distribution. These spectroscopic samples were used as part of a larger test suite to verify and characterize the performance of the four photometric redshift codes compared in this paper: ANNZ2, BPZ, SKYNET, and TPZ.

We identified challenges in producing photometric redshifts with the spectroscopic samples available to us and DES photometry, including learning the radial profile of the spectroscopic distributions in machine learning codes and mischaracterization of the redshift in template-based approaches due to the limitations of our photometric bands and template color coverage. This can result in artificial features in the photometric $n(z)$, which will bias any resulting analysis that depends on the photometric redshift distribution. We also discussed the challenge of compiling representative and complete spectroscopic training sets. However, we demonstrated that the potential bias in mean redshift due to spectroscopic incompleteness does not exceed the expected sample variance uncertainty in our presently available samples due to their small size.

In order to mitigate the potential issues associated with any given photometric redshift approach, we applied three independent methodologies: the first is based on empirical spectroscopic data and utilizing machine learning techniques; the second is a modeling-based approach, comprising a template-fitting routine (BPZ) and a first-order correction of the associated model biases by image simulations (using BCC and UFig); and finally we employed highly accurate empirical photometric redshifts from COSMOS, which have been selected to mimic our weak lensing sample. We found the mean redshift of the shear catalogue to be $z = 0.72$. The variance in this mean and those of the three tomographic bins are consistent with a Gaussian distribution of width 0.05. Therefore in the companion cosmology paper [20], we marginalize over the photometric redshift calibration uncertainty using

independent Gaussian priors of width 0.05 in each photometric redshift bin.

We propagated these photo-z uncertainties and biases through to measurements that are most relevant to weak lensing science, which is a necessary step to provide useful characterizations of photo-z biases for DES SV analysis papers. For each of the independent and deep weighted spectroscopic validation sets, we compared for each photo-z estimate the resulting measures of ξ_+ and the resulting constraints on σ_8, as well as resulting measurements of $\langle\Sigma_{\rm crit}^{-1}\rangle$. This provided us with direct estimates of expected biases on typical weak lensing measurements and cosmological parameters of interest, and allowed us to validate methods of marginalizing over photo-z biases.

We found that compared to the weighted spectroscopic validation sets, we should expect a level of bias for the fiducial photo-z estimates of less than about 10% in ξ_+, which corresponds to a 1σ deviation or bias of 2–3% in σ_8 for the fiducial SKYNET method, given DES SV statistical power. We verified an approach to mitigate this bias by marginalizing over bias parameters that shift the mean redshift of each tomographic bin, demonstrating that this is a sufficient approach to remove any bias in ξ_+ and σ_8. A similar analysis of $\langle\Sigma_{\rm crit}^{-1}\rangle$ found a bias for the fiducial photo-z estimate that increases to approximately 5% for the highest redshift lenses, but which is negligible for most lens redshifts.

Looking towards the future of the DES and beyond, weak lensing-oriented photo-z estimation will face a number of challenges. First, in order to remain comparable to the expected statistical uncertainties in a 5000 deg^2 survey, the systematic uncertainties on the mean redshift within a given tomographic bin will need to be reduced from $\delta z \sim 0.05$ to an eventual level of $\delta z \sim 0.003$. Moreover, extracting the greatest amount of the information in the lensing signal will require the use of finer tomographic binning. Finally, the detailed topology of the $p(z)$ in a given tomographic bin will come under increasing scrutiny and marginalizing over simple redshift bias parameters in the mean is unlikely to be sufficient in future cosmology analyses. Our testing metrics will need to be expanded to include those more sensitive to PDF information on a galaxy-by-galaxy basis (e.g. Ref. [11]) in order to account for this shift in emphasis.

The methodologies employed to produce photo-z's can be improved upon by exploring better galaxy templates in modeling approaches to mitigate problems observed in this work, and the incorporation of galaxy information beyond magnitude and color may be key to breaking degeneracies in the machine learning PDFs. Coupled with algorithmic improvements is the increasing availability of data. For instance, the year 1 DES survey data cover further key spectroscopic fields in Stripe 82, BOSS, DEEP2 and Wigglez. Wide field spectroscopic fields, even those biased towards the brightest objects, open up new possibilities in the form of cross-correlation analyses [64]. Meanwhile, further exquisite *photometric* fields will also be covered and should allow us to conduct comparisons similar to the one we performed with COSMOS in this work, but with reduced sample variance concerns. Despite these foreseen advances in weak lensing photo-z techniques, there still remains the separate issue of *validating* the derived redshifts. To be fully confident in both the redshifts and the estimated uncertainties that we find with the various photo-z techniques, the need for additional deep, but highly complete, spectroscopy is unavoidable.

ACKNOWLEDGMENTS

We are grateful for the extraordinary contributions of our CTIO colleagues and the DECam Construction, Commissioning and Science Verification teams in achieving the excellent instrument and telescope conditions that have made this work possible. The success of this project also relies critically on the expertise and dedication of the DES Data Management group. M. T., S. B., N. M., and J. Z. acknowledge support from the European Research Council in the form of a Starting Grant with number 240672. D. G. acknowledges the support by the DFG Cluster of Excellence Origin and Structure of the Universe. Funding for the DES Projects has been provided by the U.S. Department of Energy, the U.S. National Science Foundation, the Ministry of Science and Education of Spain, the Science and Technology Facilities Council of the United Kingdom, the Higher Education Funding Council for England, the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign, the Kavli Institute of Cosmological Physics at the University of Chicago, the Center for Cosmology and Astro-Particle Physics at the Ohio State University, the Mitchell Institute for Fundamental Physics and Astronomy at Texas A&M University, Financiadora de Estudos e Projetos, Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro, Conselho Nacional de Desenvolvimento Científico e Tecnológico and the Ministério da Ciência e Tecnologia, the Deutsche Forschungsgemeinschaft and the Collaborating Institutions in the Dark Energy Survey. C. G. acknowledges the support by the DFG Cluster of Excellence Origin and Structure of the Universe. The DES data management system is supported by the National Science Foundation under Grant Number AST-1138766. The DES participants from Spanish institutions are partially supported by MINECO under grants AYA2012-39559, ESP2013-48274, FPA2013-47986, and Centro de Excelencia Severo Ochoa SEV-2012-0234, some of which include ERDF funds from the European Union. The Collaborating Institutions are Argonne National Laboratory, the University of California at Santa Cruz, the University of Cambridge, Centro de Investigaciones Energeticas, Medioambientales y Tecnologicas-Madrid, the University of Chicago, University College London, the DES-Brazil Consortium, the Eidgenössische Technische Hochschule (ETH) Zürich, Fermi National Accelerator Laboratory, the University of Edinburgh, the University of

Illinois at Urbana-Champaign, the Institut de Ciències de l'Espai (IEEC/CSIC), the Institut de Física d'Altes Energies, Lawrence Berkeley National Laboratory, the Ludwig-Maximilians Universität and the associated Excellence Cluster Universe, the University of Michigan, the National Optical Astronomy Observatory, the University of Nottingham, The Ohio State University, the University of Pennsylvania, the University of Portsmouth, SLAC National Accelerator Laboratory, Stanford University, the University of Sussex, and Texas A&M University. Based in part on observations taken at the Australian Astronomical Observatory under program A/2013B/012. Parts of this research were conducted by the Australian Research Council Centre of Excellence for All-sky Astrophysics (CAASTRO), through project number CE110001020. This work was supported in part by grants 200021_14944 and 200021_143906 from the Swiss National Science Foundation. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 179.A-2004. Based on observations made with ESO Telescopes at the La Silla Paranal Observatory under programme ID 177.A-3016. This paper is Fermilab publication FERMILAB-PUB-15-306 and DES publication DES2015-0060. This paper has gone through internal review by the DES Collaboration.

APPENDIX A: DETAILS OF MATCHED SPECTROSCOPIC SAMPLE

In this appendix, we note all the quality flags that are used in the matched spectroscopic catalogue and their meaning.

(i) *2dFGRS*: All galaxies with flags 3, 4 or 5; all of these are considered to be reliable redshifts [65].
(ii) *ACES*: All galaxies with flags 3 or 4; these are labeled as secure and very secure redshifts [66].
(iii) *ATLAS*: This survey [67] has no quality flags; all objects classified as galaxies where kept.
(iv) *OzDES*: All galaxies with quality flag 4; galaxies with this flag are expected to have the correct redshift more than 99% of the time [68].
(v) *ELG Cosmos*: All galaxies with quality flags 3 or 4; these correspond to clear single line redshift identification and a secure redshift respectively [69].
(vi) *GAMA*: All galaxies with quality flag 4; these are labeled as certain redshifts [70].
(vii) *PanSTARRS AAOmega*: All galaxies with quality flag 3 or 4; galaxies with these flags are expected to have the correct redshift more than 95% or 99% of the time, respectively. [71–73].
(viii) *PanSTARRS MMT*: All galaxies with quality flag 3 or 4; these are labeled as probably and as certain redshifts [71–73].
(ix) *SDSS DR10*: All galaxies with quality flag 0; this are labeled as reliable [74].
(x) *SNLS AAOmega*: All galaxies with quality flag 4 and 5; these are labeled as reliable and reliable with more the three clearly visible features [75].
(xi) *SNLS*: All galaxies with quality flag 1 and 2; these are labeled as reliable based on several strong detected features and on one clearly detected feature, usually [OII] [76].
(xii) *UDS*: All galaxies observed with VIMOS that have quality flags 3 and 4; these are labeled as secure. All galaxies observed with FORS2 that have quality flags A, B or B* where A and B is labeled as secure and B* is labeled as reliable. See Refs. [77,78] for more information.
(xiii) *VIPERS*: All galaxies that have flags 3 and 4; these are labeled as reliable [79].
(xiv) *Zcosmos*: All galaxies that have flags 3 and 4; these are labeled as secure and very secure redshifts [80].
(xv) *VVDS*: All galaxies that have flags 3 and 4, these are labeled as secure and very secure redshifts [81,82].

APPENDIX B: SELF-SELECTION TOMOGRAPHIC ANALYSIS

We repeat here Tests 1, 2a, and 2b, but now allow each code to assign a galaxy to each redshift bin based on its own estimate of the mean PDF instead of that of SKYNET as was done in Sec. V. Figures 17, 18, and 19 show the performance of the four methods. Table VI shows the offsets of the mean of the redshift estimated distributions with respect to the weighted spectroscopic distribution. There is not a clear benefit to enforcing separate tomographic binning based on each photo-z method and repeating the analysis pipelines in the companion papers for DES SV, as some methods perform better and others worse when using the fiducial SKYNET binning.

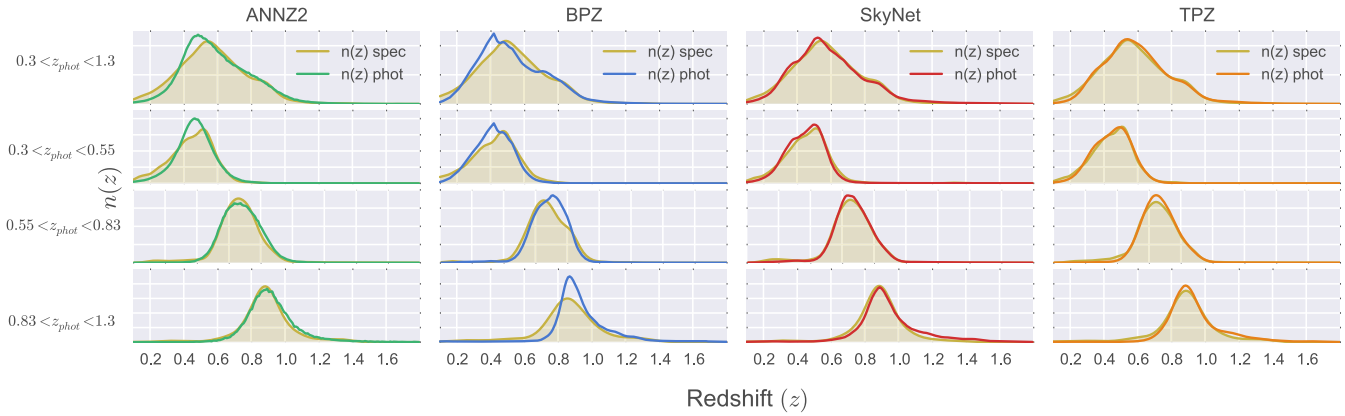

FIG. 17. The weighted spectroscopic redshift distribution $n(z)$ (shaded area) compared to the estimates of the four codes for the Test 1 (VVDS-F14) galaxies. Unlike in Sec. V, all codes assign galaxies to tomographic bins according to their own mean PDF estimates, and hence the objects in each bin differ for each panel in the plot.

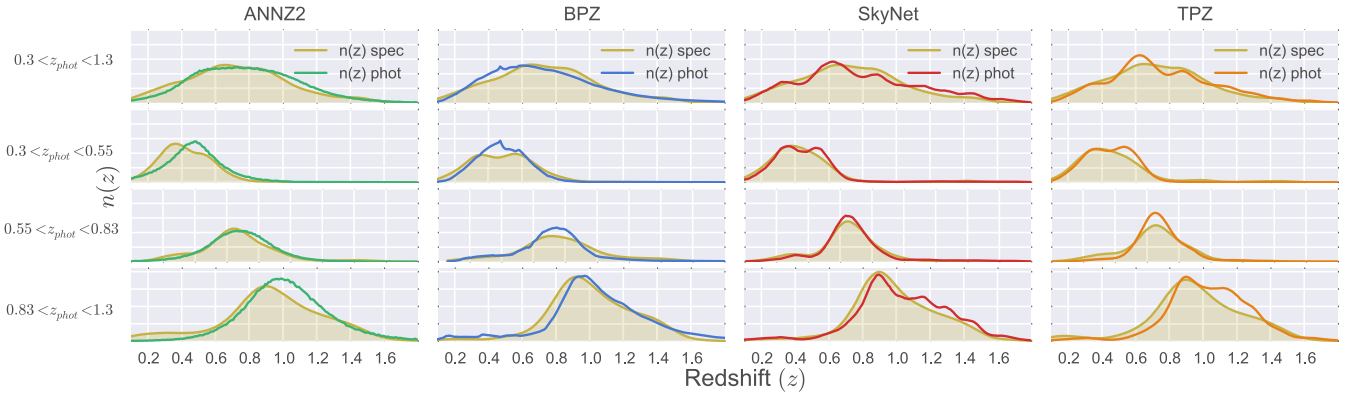

FIG. 18. The weighted spectroscopic redshift distribution $n(z)$ (shaded area) compared to the estimates of the four codes in Test 2a (VVDS-Deep galaxies in the validation set). Unlike in Sec. V, all codes assign galaxies to tomographic bins according to their own mean PDF estimates, and hence the objects in each bin differ for each panel in the plot.

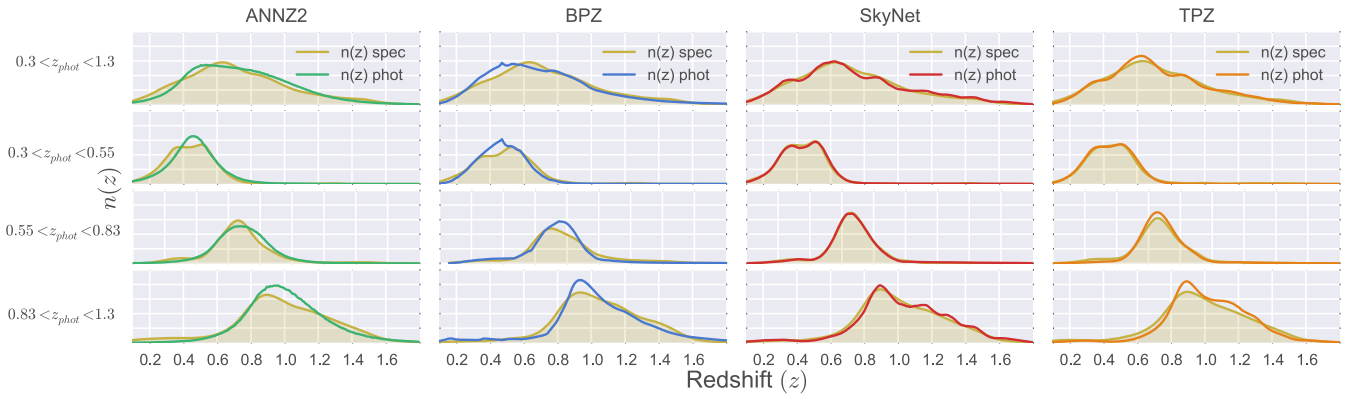

FIG. 19. The weighted spectroscopic redshift distribution $n(z)$ (shaded area) compared to the estimates of the four codes in Test 2b (full validation set). Unlike in Sec. V, all codes assign galaxies to tomographic bins according to their own mean PDF estimates, and hence the objects in each bin differ for each panel in the plot.

TABLE VI. The bias $(\langle z_{\rm phot}\rangle - \langle z_{\rm spec}\rangle)$ between the photometric redshift estimates and the true spectroscopic distribution in Test 1 (independent), Test 2a (VVDS-Deep), and Test 2b (full validation set) when the codes each assign their own binning to the galaxies.

	z range	ANNZ2	BPZ	SKYNET	TPZ
Test 1	0.30–0.55	0.017	−0.005	0.003	0.004
	0.55–0.83	0.018	0.01	0.017	0.016
	0.83–1.30	0.032	0.077	0.063	0.050
Test 2a	0.30–0.55	0.049	0.002	0.027	−0.013
	0.55–0.83	0.015	−0.025	0.034	0.031
	0.83–1.30	0.086	0.046	0.044	0.069
Test 2b	0.30–0.55	0.015	−0.015	0.012	−0.020
	0.55–0.83	0.011	−0.027	0.013	0.008
	0.83–1.30	0.025	0.007	0.022	0.028

APPENDIX C: PHOTO-Z METHODS

1. ANNZ2

ANNZ2 [83][4] is an updated version of the neural network code ANNZ ([84]). ANNZ2 differs from its previous version by incorporating several additional machine learning methods beyond artificial neural networks (ANNs), such as boosted decision trees and k-nearest neighbors (KNN) algorithms. These are implemented in the TMVA package [85].[5]

For the 100 ANNs run on the spectroscopic training set, we randomly varied the number of nodes in each layer, the number of training cycles, the usage of the so-called *Bayesian regulator*, that reduces the risk of over-training, the type of activation function, the type of variable transformation performed before training (such as normalization and principal component analysis transformation), the number of subsequent convergence tests which have to fail to consider the training complete, and the initial random seed. After training is complete, the performance of each method is quantified through an optimization process, which leads to a single nominal photo-z estimator for ANNZ2. The entire collection of solutions is used in order to derive a $p(z)$, constructed in two steps. First, each solution is folded with an error distribution, which is derived using the KNN error estimation method of Ref. [87]. The ensemble of solutions is then combined using an optimized weighting scheme. This methodology allows us to take into account both the intrinsic errors on the input parameters for a given method, and the uncertainty on the method itself. The methodology described above is what is called "randomized regression." Another important feature implemented in ANNZ2 is the weighting method [59]. It is therefore possible to give as input a reference sample and reweight the training set to make its relevant variable distributions more representative of the former; this technique was applied in this work.

2. BPZ

The BPZ (Bayesian photometric redshifts) photo-z code [60,61] is a model-fitting code that fits galaxy templates to the measured photometry and its associated errors. BPZ calculates the likelihood of the galaxy for the best-fitting template, which then, using Bayes theorem, is combined with a prior to produce the likelihood. The prior represents our previous knowledge of the redshift and spectral type distributions of the sample in the analysis.

(i) Templates: We use the eight spectral templates that BPZ carries by default based on Refs. [88,89], and add two more interpolated templates between each pair of them by setting the input parameter INTERP=8 (option by default).

(ii) Prior: We explicitly calibrate the prior in each test by fitting the empirical function $\Pi(z, t|m_0)$ proposed in Ref. [60] to the weighted training set, although we note that using the weighted or unweighted training set to get the prior had a negligible effect on photo-z performance.

3. SkyNet

SKYNET [90] is a neural network algorithm that uses a second-order method based on a conjugate gradient algorithm to find the optimal weights of the network. SKYNET classifies galaxies in classes, in this case redshift bins, where the last layer is a softmax transformation that is able to estimate the probability that an object belongs to a certain class (or bin) [91]. The number of classes is the redshift bin resolution of the PDF. In this work SKYNET is run slightly different than in Refs. [26,91]. SKYNET is run ten times with the same network configuration but with a slightly shifted binning each time. We train with a nominal bin width of $\Delta z = 0.09$; these are referred to as the broad bins. The broad bins are then slightly shifted by $\delta = 0.009$ every training run so that Δz is sampled in ten locations, leading to an overall sampling of $\delta z = 0.009$. This produces 200 bins between $z = 0.005$ and $z = 1.8$. After the ten networks have been trained, the PDF values at z_i are taken to be the average of all the broad bins that z_i lies within. This means that the SKYNET photometric redshifts have an intrinsic smoothing built into them. All the networks have the same architecture, three layers with 16, 14, and 20 nodes per layer and a *tanh* activation function. The features fed to the network are the MAG_AUTO i, r and all possible color combinations of the four bands. In this work we make use of the PYTHON wrapper PYSKYNET[6] of the SKYNET library.

[4] https://github.com/IftachSadeh/ANNZ.

[5] TMVA is a part of the ROOT C++ software framework [86].

[6] http://pyskynet.readthedocs.org/.

4. TPZ

TPZ[7] [92] is a machine learning algorithm that uses prediction trees and random forest techniques to produce robust photometric redshift PDFs. Prediction trees are built by asking a series of questions that recursively split the input data taken from the spectroscopic sample into two branches, until a terminal leaf is created that meets the stopping criterion. The method by which the data are divided is chosen to be the one with the highest information gain among the random subsample of features chosen at every point. This produces less correlated trees that act as weak learners that can be combined into a strong predictor. All objects in a terminal leaf node represent a specific subsample of the entire data with similar properties. Additional data is created before the trees are constructed by perturbing the data using their magnitude errors; this is sometimes referred to as a parametric bootstrap. In this work 200 trees were created whose results were aggregated to construct each individual PDF. For the application to DES SV data, we have used *griz* MAG_AUTO magnitudes together with all the corresponding colors and their associated errors. We discretized the redshift space into 100 bins up to $z = 1.8$ and adopted a smoothing scale of 5 times the bin size.

APPENDIX D: PRIMUS, AN EXAMPLE OF EXTREME SELECTION EFFECTS

In building the spectroscopic training and validation samples, we have excluded any galaxies from the PRIMUS survey [93]. Here we will discuss some of the complications of using PRIMUS galaxies as part of the training or validation samples. PRIMUS is a spectroscopic survey covering a total of 9.1 deg^2 containing 185 105 galaxies, of which we have matched 88 040 galaxies that have DES SV photometry within 1.5 arcseconds only using the two highest PRIMUS quality flags 4 and 3. The PRIMUS redshifts are obtained by fitting low-resolution spectra and any matched photometry to an empirical library of spectra based on the AGES spectra [94]. The PRIMUS redshifts have two peculiarities, the first being that a non-negligible amount of galaxies have a different redshift when compared to objects with spectra from higher-resolution instruments. The authors of Ref. [93] estimated $\sigma_{\delta z/(1+z)} = 0.005$ and 0.022 for quality flags 4 and 3, while we find 0.004 and 0.010 for all the matched objects within the DES survey. The top two panels of Fig. 20 show this comparison of the PRIMUS spectroscopic redshifts with matched spectroscopic redshifts from higher-resolution instruments. This leads us to consider the unresolved question of how robust ML and other calibration methods are to incorrect spectra, which is not a question that we attempt to answer in this work, but one for which there has been some work in general in the ML literature (e.g., Refs. [15,95]).

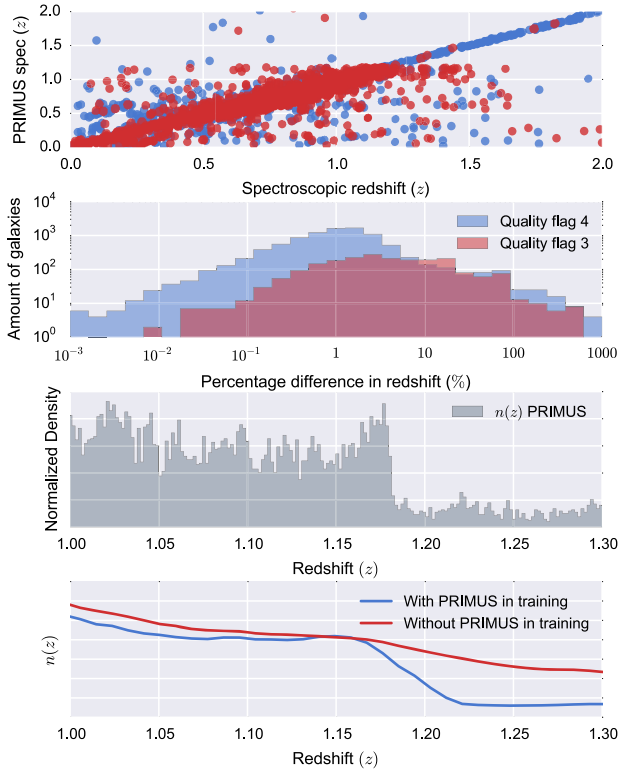

FIG. 20. An analysis of challenges related to the use of PRIMUS spectroscopic redshifts as part of the DES SV training or validation samples. Top panel: PRIMUS redshift vs the matched spectroscopic redshift from higher-resolution instruments. The blue dots are the highest quality flag 4, while the red dots are the second highest quality flag 3. Second panel: The fractional difference of the redshifts between PRIMUS and the other surveys. Third panel: The spectroscopic redshift distribution of PRIMUS galaxies between $1.0 < z < 1.3$. Around $\sim$1.2, there is a large drop in the spectroscopic redshift distribution due to the fact that the galaxies have a maximum fitting redshift of $z = 1.2$, while AGN are fit up to $z = 5.0$. Bottom panel: The effect of this drop in PRIMUS $n(z)$ on the final estimation of the $n(z)$ for the DES SV shear catalogue. Shown are two examples of including or excluding the PRIMUS galaxies using SKYNET, where the feature at $z = 1.2$ is clearly imprinted on the $n(z)$ of the weak lensing sample when PRIMUS galaxies are included in the training.

The second PRIMUS feature that is important for photometric redshift estimation is the fact that galaxies are only fit up to $z = 1.2$. The cut at $z = 1.2$ is effectively a selection effect and hence, one must take care when using PRIMUS to train. To illustrate this, consider a galaxy at $z = 1.2$ observed by PRIMUS and DES for which we want to estimate the $p(z)$. In the idealized case of a Gaussian PDF, the mean would be located around $z = 1.2$ and there would be tails in the $p(z)$ extending to lower and higher redshift. Given that there are no galaxies beyond $z = 1.2$ in PRIMUS, none of the ML methods will be able to learn that

[7] http://lcdm.astro.illinois.edu/research/TPZ.html.

some probability should extend beyond $z = 1.2$. Even when assessing how well a template-fitting method performs, the lack of spectra beyond $z = 1.2$ may lead one to believe the performance is poor. These features are demonstrated in the bottom two panels of Fig. 20. In the bottom panel of Fig. 20, we provide a real example of the difference on the reconstructed $n(z)$ for the weak lensing sample around $z = 1.2$ when trained with and without PRIMUS. Though this is an extreme case of a selection effect imprinting itself on the reconstructed $n(z)$, it is possible that similar, more subtle effects persist in the ML photometric redshift estimates. There are a large number of PRIMUS spectra, however, and careful efforts should be made to find ways to utilize these in the future.

[1] D. Munshi, P. Valageas, L. van Waerbeke, and A. Heavens, Phys. Rep. **462**, 67 (2008).
[2] R. Narayan and M. Bartelmann, in *Formation of Structure in the Universe*, edited by A. Dekel and J. P. Ostriker (Cambridge University Press, Cambridge, 1999), p. 360.
[3] A. Refregier, Annu. Rev. Astron. Astrophys. **41**, 645 (2003).
[4] M. Jarvis *et al.* (to be published).
[5] W. Hu, Astrophys. J. Lett. **522**, L21 (1999).
[6] F. B. Abdalla, A. Amara, P. Capak, E. S. Cypriano, C. Lahav, and J. Rhodes, Mon. Not. R. Astron. Soc. **387**, 969 (2008).
[7] A. Amara and A. Réfrégier, Mon. Not. R. Astron. Soc. **381**, 1018 (2007).
[8] M. Banerji, F. B. Abdalla, O. Lahav, and H. Lin, Mon. Not. R. Astron. Soc. **386**, 1219 (2008).
[9] F. Bellagamba, M. Meneghetti, L. Moscardini, and M. Bolzonella, Mon. Not. R. Astron. Soc. **422**, 553 (2012).
[10] G. Bernstein and D. Huterer, Mon. Not. R. Astron. Soc. **401**, 1399 (2010).
[11] R. Bordoloi, S. J. Lilly, and A. Amara, Mon. Not. R. Astron. Soc. **406**, 881 (2010).
[12] R. Bordoloi *et al.*, Mon. Not. R. Astron. Soc. **421**, 1671 (2012).
[13] Y.-C. Cai, R. E. Angulo, C. M. Baugh, S. Cole, C. S. Frenk, and A. Jenkins, Mon. Not. R. Astron. Soc. **395**, 1185 (2009).
[14] C. E. Cunha, D. Huterer, M. T. Busha, and R. H. Wechsler, Mon. Not. R. Astron. Soc. **423**, 909 (2012).
[15] C. E. Cunha, D. Huterer, H. Lin, M. T. Busha, and R. H. Wechsler, Mon. Not. R. Astron. Soc. **444**, 129 (2014).
[16] E. S. Sheldon, C. E. Cunha, R. Mandelbaum, J. Brinkmann, and B. A. Weaver, Astrophys. J. Suppl. Ser. **201**, 32 (2012).
[17] L. Sun, Z.-H. Fan, C. Tao, J. P. Kneib, S. Jouvel, and A. Tilquin, Astrophys. J. **699**, 958 (2009).
[18] D. Kirk *et al.*, Space Sci. Rev. **193**, 139 (2015).
[19] M. A. Troxel and M. Ishak, Phys. Rev. **558**, 1 (2015).
[20] T. Abbott *et al.* (The Dark Energy Survey Collaboration, arXiv:1507.05552.
[21] H. Hildebrandt *et al.*, Astron. Astrophys. **523**, A31 (2010).
[22] M. Banerji *et al.*, Mon. Not. R. Astron. Soc. **446**, 2523 (2015).
[23] J. Benjamin *et al.*, Mon. Not. R. Astron. Soc. **431**, 1547 (2013).
[24] H. Hildebrandt *et al.*, Mon. Not. R. Astron. Soc. **421**, 2355 (2012).
[25] R. Mandelbaum *et al.*, Mon. Not. R. Astron. Soc. **386**, 781 (2008).
[26] C. Sánchez *et al.*, Mon. Not. R. Astron. Soc. **445**, 1482 (2014).
[27] S. J. Schmidt and P. Thorman, Mon. Not. R. Astron. Soc. **431**, 2766 (2013).
[28] T. Erben *et al.*, Mon. Not. R. Astron. Soc. **433**, 2545 (2013).
[29] C. Heymans *et al.*, Mon. Not. R. Astron. Soc. **432**, 2433 (2013).
[30] D. M. Wittman *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **4836**, 73 (2002).
[31] H. Diehl, Phys. Procedia **37**, 1332 (2012).
[32] B. Flaugher *et al.*, Astron. J. **150**, 150 (2015).
[33] B. L. Flaugher *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **8446**, 844611 (2012).
[34] K. Honscheid *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **8451**, 845112 (2012).
[35] B. Leistedt *et al.* (to be published).
[36] J. E. Carlstrom *et al.*, Publ. Astron. Soc. Pac. **123**, 568 (2011).
[37] J. J. Mohr *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **8451**, 84510D (2012).
[38] E. Bertin, in *Astronomical Society of the Pacific Conference Series, Vol. 351, Astronomical Data Analysis Software and Systems XV*, edited by C. Gabriel, C. Arviset, D. Ponz, and S. Enrique (Astronomical Society of the Pacific, San Francisco, 2006), p. 112.
[39] E. Bertin, Y. Mellier, M. Radovich, G. Missonnier, P. Didelon, and B. Morin, in *Astronomical Society of the Pacific Conference Series, Vol. 281, Astronomical Data Analysis Software and Systems XI*, edited by D. A. Bohlender, D. Durand, and T. H. Handley (Astronomical Society of the Pacific, San Francisco, 2002), p. 228.
[40] S. Desai *et al.*, Astrophys. J. **757**, 83 (2012).
[41] I. Sevilla *et al.*, arXiv:1109.6741.
[42] E. Bertin, in *Astronomical Society of the Pacific Conference Series, Vol. 442, Astronomical Data Analysis Software and Systems XX*, edited by I. N. Evans, A. Accomazzi, D. J. Mink, and A. H. Rots (Astronomical Society of the Pacific, San Francisco, 2011), p. 435.
[43] E. Bertin and S. Arnouts, Astron. Astrophys. Suppl. Ser. **117**, 393 (1996).
[44] M. Crocce *et al.* (to be published).
[45] E. Rykoff *et al.* (to be published).
[46] J. Zuntz, T. Kacprzak, L. Voigt, M. Hirsch, B. Rowe, and S. Bridle, Mon. Not. R. Astron. Soc. **434**, 1604 (2013).

[47] D. W. Hogg and D. Lang, Publ. Astron. Soc. Pac. **125**, 719 (2013).
[48] L. Miller, T. D. Kitching, C. Heymans, A. F. Heavens, and L. van Waerbeke, Mon. Not. R. Astron. Soc. **382**, 315 (2007).
[49] R. Mandelbaum *et al.*, Astrophys. J. Suppl. Ser. **212**, 5 (2014).
[50] Ž. Ivezić, A. Connolly, J. Vanderplas, and A. Gray, *Statistics, Data Mining and Machine Learning in Astronomy* (Princeton University Press, Princeton, NJ, 2014).
[51] O. Ilbert *et al.*, Astrophys. J. **690**, 1236 (2009).
[52] S. Arnouts, S. Cristiani, L. Moscardini, S. Matarrese, F. Lucchin, A. Fontana, and E. Giallongo, Mon. Not. R. Astron. Soc. **310**, 540 (1999).
[53] C. Bruderer *et al.*, Astrophys. J. **817**, 25 (2016).
[54] J. Bergé, L. Gamper, A. Réfrégier, and A. Amara, Astronomical Computing, **1**, 23 (2013).
[55] M. T. Busha, R. H. Wechsler, M. R. Becker, B. Erickson, and A. E. Evrard, American Astronomical Society Meeting Abstracts **221**, 341.07 (2013).
[56] T. Kacprzak, S. Bridle, B. Rowe, L. Voigt, J. Zuntz, M. Hirsch, and N. MacCrann, Mon. Not. R. Astron. Soc. **441**, 2528 (2014).
[57] L. M. Voigt and S. L. Bridle, Mon. Not. R. Astron. Soc. **404**, 458 (2010).
[58] E. S. Rykoff, E. Rozo, and R. Keisler, arXiv:1509.00870.
[59] M. Lima, C. E. Cunha, H. Oyaizu, J. Frieman, H. Lin, and E. S. Sheldon, Mon. Not. R. Astron. Soc. **390**, 118 (2008).
[60] N. Benítez, Astrophys. J. **536**, 571 (2000).
[61] D. Coe, N. Benítez, S. F. Sánchez, M. Jee, R. Bouwens, and H. Ford, Astron. J. **132**, 926 (2006).
[62] M. R. Becker *et al.*, Phys. Rev. D **94**, 022002 (2016).
[63] E. S. Rykoff, E. Rozo, and R. Keisler, Mon. Not. R. Astron. Soc. **461**, 1431 (2016).
[64] J. A. Newman, Astrophys. J. **684**, 88 (2008).
[65] M. Colless *et al.*, Mon. Not. R. Astron. Soc. **328**, 1039 (2001).
[66] M. C. Cooper *et al.*, Mon. Not. R. Astron. Soc. **425**, 2116 (2012).
[67] M. Y. Mao *et al.*, Mon. Not. R. Astron. Soc. **426**, 3334 (2012).
[68] F. Yuan *et al.*, The Astronomer's Telegram **8464** (2015).
[69] J. Comparat *et al.*, Astron. Astrophys. **575**, A40 (2015).
[70] S. P. Driver *et al.*, Mon. Not. R. Astron. Soc. **413**, 971 (2011).
[71] N. Kaiser *et al.*, Proc. SPIE Int. Soc. Opt. Eng. **7733**, 77330E (2010).
[72] A. Rest *et al.*, Astrophys. J. **795**, 44 (2014).
[73] D. Scolnic *et al.*, Astrophys. J. **795**, 45 (2014).
[74] C. P. Ahn *et al.*, Astrophys. J. Suppl. Ser. **211**, 17 (2014).
[75] C. Lidman *et al.*, Mon. Not. R. Astron. Soc. **427**, 550 (2012).
[76] C. Balland *et al.* (to be published).
[77] E. J. Bradshaw *et al.*, Mon. Not. R. Astron. Soc. **433**, 194 (2013).
[78] R. J. McLure *et al.*, Mon. Not. R. Astron. Soc. **428**, 1088 (2013).
[79] B. Garilli *et al.*, Astron. Astrophys. **562**, A23 (2014).
[80] S. J. Lilly *et al.*, Astrophys. J. Suppl. Ser. **184**, 218 (2009).
[81] B. Garilli *et al.*, Astron. Astrophys. **486**, 683 (2008).
[82] O. Le Fèvre *et al.*, Astron. Astrophys. **439**, 845 (2005).
[83] I. Sadeh, F. B. Abdalla, and O. Lahav, arXiv:1507.00490.
[84] A. A. Collister and O. Lahav, Publ. Astron. Soc. Pac. **116**, 345 (2004).
[85] A. Hoecker *et al.*, arXiv:physics/0703039.
[86] R. Brun and F. Rademakers, Nucl. Instrum. Methods Phys. Res., Sect. A **389**, 81 (1997).
[87] H. Oyaizu, M. Lima, C. E. Cunha, H. Lin, and J. Frieman, Astrophys. J. **689**, 709 (2008).
[88] G. D. Coleman, C.-C. Wu, and D. W. Weedman, Astrophys. J. Suppl. Ser. **43**, 393 (1980).
[89] A. L. Kinney, D. Calzetti, R. C. Bohlin, K. McQuade, T. Storchi-Bergmann, and H. R. Schmitt, Astrophys. J. **467**, 38 (1996).
[90] P. Graff and F. Feroz, SkyNet: Neural network training tool for machine learning in astronomy, http://ascl.net/1312.007.
[91] C. Bonnett, Mon. Not. R. Astron. Soc. **449**, 1043 (2015).
[92] M. Carrasco Kind and R. J. Brunner, Mon. Not. R. Astron. Soc. **432**, 1483 (2013).
[93] R. J. Cool *et al.*, Astrophys. J. **767**, 118 (2013).
[94] C. S. Kochanek *et al.*, Astrophys. J. Suppl. Ser. **200**, 8 (2012).
[95] D. F. Nettleton, A. Orriols-Puig, and A. Fornells, Artif. Intell. Rev. **33**, 275 (2010).